



RT "Molecular cloning, sequencing and characterization of cDNA to rat  
 RL liver rhodanese, a thiosulphate sulphurtransferase.";  
 RN Biochem. J. 275:227-231(1991).  
 RP [2]  
 RN MUTAGENESIS.  
 RP TISSUE=Liver;  
 RC MEDLINE=95332330; PubMed=7608189;  
 RX Nagahara N., Okazaki T., Nishino T.;  
 RA "Cytosolic mercaptopyruvate sulfurtransferase is evolutionarily  
 RT related to mitochondrial rhodanese. Striking similarity in active site  
 RT amino acid sequence and the increase in the mercaptopyruvate  
 RT sulfurtransferase activity of rhodanese by site-directed  
 RT mutagenesis.";  
 RL J. Biol. Chem. 270:16230-16235(1995).  
 CC -1- FUNCTION: INVOLVED IN THE FORMATION OF IRON-SULFUR COMPLEXES,  
 CC CYANIDE DETOXIFICATION OR MODIFICATION OF SULFUR-CONTAINING  
 CC ENZYMES. OTHER THIOL COMPOUNDS, BESIDES CYANIDE, CAN ACT AS SULFUR  
 CC ION ACCEPTORS. ALSO HAS WEAK MERCAPTOPYRUVATE SULFURTRANSFERASE  
 CC (MST) ACTIVITY.  
 CC -1- CATALYTIC ACTIVITY: Thiosulfate + cyanide = sulfite + thiocyanate.  
 CC -1- SUBUNIT: Monomer.  
 CC -1- SUBCELLULAR LOCATION: Mitochondrial matrix.  
 CC -1- TISSUE SPECIFICITY: FOUND IN NUMEROUS TISSUES.  
 CC -1- DOMAIN: THE STRUCTURE CONSISTS OF 2 DOMAINS OF VERY SIMILAR  
 CC CONFORMATION, SUGGESTING A COMMON EVOLUTIONARY ORIGIN. HOWEVER,  
 CC THE SEQUENCES OF THE 2 DOMAINS ARE VERY DIFFERENT.  
 CC -1- SIMILARITY: Contains 2 rhodanese domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X56228; CAA39677.1; --  
 DR PIR; S15081; S15081.  
 DR HSBP; P00586; 1RS8.  
 DR InterPro; IPR001763; Rhodanese-like.  
 DR InterPro; IPR001307; Rhodanese.  
 DR Pfam; PF00581; Rhodanese; 2.  
 DR SMART; SM00450; RHOD; 2.  
 DR PROSITE; PS00380; RHODANESE 1; 1.  
 DR PROSITE; PS00683; RHODANESE 2; 1.  
 DR PROSITE; PSS0206; RHODANESE 3; 2.  
 KW Transferase; Mitochondrion; Repeat.  
 FT NON\_TER 1  
 FT DOMAIN 23 141  
 FT DOMAIN 142 157  
 FT DOMAIN 171 286  
 FT ACT\_SITE 185 185  
 FT ACT\_SITE 246 246  
 FT ACT\_SITE 247 247  
 FT ACT\_SITE 248 248  
 FT MUTAGEN 247 247  
 FT MUTAGEN 248 248  
 FT SEQUENCE 295 AA; 33176 MW; 24C55B35690934E1 CRC64;  
 Query Match 84.4%; Score 27; DB 1; Length 295;  
 Best Local Similarity 71.4%; Pred. No. 44;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ATKNSL 7  
 DB 153 ATLNRSL 159  
 RESULT 15  
 THTR\_BOVIN

THTR\_BOVIN STANDARD; PRT; 296 AA.  
 P00586;  
 21-JUL-1986 (Rel. 01, Created)  
 01-AUG-1992 (Rel. 23, Last sequence update)  
 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Thiosulfate sulfurtransferase (EC 2.8.1.1) (Rhodanese).  
 GN TST.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 [1]  
 SEQUENCE FROM N.A.  
 RP MEDLINE=91161544; PubMed=2002017;  
 RA Miller D.M., Delgado R., Chirgwin J.M., Hardies S.C., Horowitz P.M.;  
 RT "Expression of cloned bovine adrenal rhodanese.";  
 RL J. Biol. Chem. 266:4686-4691(1991).  
 [2]  
 SEQUENCE OF 1-294.  
 RP TISSUE=Liver;  
 RC MEDLINE=79048424; PubMed=711737;  
 RX Russell J., Weng L., Keim P.S., Heinrikson R.L.;  
 RA "The covalent structure of bovine liver rhodanese. Isolation and  
 RT partial structural analysis of cyanogen bromide fragments and the  
 RT complete sequence of the enzyme.";  
 RL J. Biol. Chem. 253:8102-8108(1978).  
 [3]  
 X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).  
 RP MEDLINE=79007483; PubMed=691057;  
 RA Ploegman J.H., Drent G., Kalk K.H., Hol W.G.J.;  
 RT "Structure of bovine liver rhodanese. I. Structure determination at  
 RT 2.5-A resolution and a comparison of the conformation and sequence of  
 RT its two domains.";  
 RL J. Mol. Biol. 123:557-594(1978).  
 [4]  
 X-RAY CRYSTALLOGRAPHY (1.36 ANGSTROMS).  
 RP MEDLINE=98437562; PubMed=9761843;  
 RA Gliubich F., Berni R., Colapietro M., Barba L., Zanotti G.;  
 RT "Structure of sulfur-substituted rhodanese at 1.36-A resolution.";  
 RL Acta Crystallogr. D 54:481-486(1998).  
 [5]  
 ACTIVE SITE.  
 RP MEDLINE=79048425; PubMed=711738;  
 RX Weng L., Heinrikson R.L., Westley J.;  
 RA "Active site cysteinyl and arginyl residues of rhodanese. A novel  
 RT formation of disulfide bonds in the active site promoted by  
 RT phenylethoxal.";  
 RL J. Biol. Chem. 253:8109-8119(1978).  
 [6]  
 MUTAGENESIS OF ARG-186 AND LYS-249.  
 RP MEDLINE=94179198; PubMed=8132546;  
 RX Luo G.-X., Horowitz P.M.;  
 RT "The sulfurtransferase activity and structure of rhodanese are  
 RT affected by site-directed replacement of Arg-186 or Lys-249.";  
 RL J. Biol. Chem. 269:8220-8225(1994).  
 CC -1- FUNCTION: FORMATION OF IRON-SULFUR COMPLEXES AND CYANIDE  
 CC DETOXIFICATION. BINDS MOLECULAR OXYGEN AND SULFUR.  
 CC -1- CATALYTIC ACTIVITY: Thiosulfate + cyanide = sulfite + thiocyanate.  
 CC -1- SUBUNIT: Monomer.  
 CC -1- SUBCELLULAR LOCATION: Mitochondrial matrix..  
 CC -1- TISSUE SPECIFICITY: FOUND IN NUMEROUS TISSUES.  
 CC -1- DOMAIN: THE STRUCTURE CONSISTS OF 2 DOMAINS OF VERY SIMILAR  
 CC CONFORMATION, SUGGESTING A COMMON EVOLUTIONARY ORIGIN. HOWEVER,  
 CC THE SEQUENCES OF THE 2 DOMAINS ARE VERY DIFFERENT.  
 CC -1- SIMILARITY: Contains 2 rhodanese domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 16:55:29 ; Search time 9.33333 Seconds  
(without alignments)  
72.127 Million cell updates/sec

Title: US-09-901-187C-3

Perfect score: 32

Sequence: 1 ATINKSL 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 76.\*

1: Pirl.\*

2: Pirl.\*

3: Pirl.\*

4: Pirl.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	32	100.0	118	2 C41298	kinesin-like prote
2	29	90.6	557	2 S62002	hypothetical prote
3	29	90.6	1516	2 T41235	probable myosin he
4	28	87.5	368	1 Q0BEHG	early nuclear anti
5	28	87.5	403	2 T09322	DNA polymerase pro
6	28	87.5	439	2 T49189	kinesin heavy chai
7	28	87.5	519	2 S66673	disulfide isomeras
8	28	87.5	550	2 T32543	hypothetical prote
9	28	87.5	636	2 H75635	hypothetical prote
10	28	87.5	779	2 A11602	translation initia
11	28	87.5	782	2 A11602	translation initia
12	28	87.5	947	2 S08243	beta-glucosidase (
13	28	87.5	1556	2 S59393	probable membrane
14	28	87.5	1588	2 T38660	probable transcrip
15	27	84.4	225	2 A2140	hypothetical prote
16	27	84.4	295	2 S15081	thiosulfate sulfur
17	27	84.4	297	1 ROBO	x-Pro aminopeptida
18	27	84.4	437	2 AG0111	hypothetical prote
19	27	84.4	491	2 T19505	probable RNA helic
20	27	84.4	506	1 S31229	medium-chain-fatty
21	27	84.4	507	2 A93458	kinesin-like prote
22	27	84.4	883	2 T40128	kinesin-like prote
23	27	84.4	1150	1 A55289	competence factor
24	27	84.4	1208	2 T09049	kinesin-related pr
25	27	84.4	1584	1 JN0114	kinesin-like prote
26	27	84.4	1584	2 T15822	kinesin family pro
27	27	84.4	1695	2 A56921	hypothetical prote
28	26	81.2	121	2 A71730	sarcocystatin A pr
29	26	81.2	122	2 A43644	

30	26	81.2	143	2 G86698	transcription regu
31	26	81.2	158	2 T30791	hypothetical prote
32	26	81.2	158	2 B42508	F15L protein - vac
33	26	81.2	161	2 A36841	C19L protein - var
34	26	81.2	161	2 E72155	F15L protein - var
35	26	81.2	161	2 T28477	hypothetical prote
36	26	81.2	173	2 T42235	FKHR protein homol
37	26	81.2	174	2 A83988	hypothetical prote
38	26	81.2	175	2 AD1327	cell-division init
39	26	81.2	175	2 AF1698	cell-division init
40	26	81.2	234	2 A51574	glycerol uptake fa
41	26	81.2	234	2 AG1220	probable carnitine
42	26	81.2	238	2 E71406	hexaprenyldihydrox
43	26	81.2	248	2 AG3275	3-demethylubiquino
44	26	81.2	250	2 AC2988	dihydroxypolypreny
45	26	81.2	250	2 E98295	

ALIGNMENTS

RESULT 1

C41298  
kinesin-like protein 3 - fruit fly (Drosophila melanogaster) (fragment)  
C:Species: Drosophila melanogaster  
C>Date: 17-Jul-1992 #sequence\_revision 17-Jul-1992 #text\_change 24-Sep-1999  
C:Accession: C41298  
R:Stewart, R.J.; Pesavento, P.A.; Woerpel, D.N.; Goldstein, L.S.B.  
Proc. Natl. Acad. Sci. U.S.A. 88, 8470-8474, 1991  
A:Title: Identification and partial characterization of six members of the kinesin superfamily: ATP; P-loop  
A:Reference number: A41298; MUID:92020874; PMID:1924306  
A:Accession: C41298  
A:Molecule type: DNA  
A:Residues: 1-118 <STE>  
A:Cross-references: GB:M74429; NID:g157787; PIDN:AAA28656.1; PID:g157788  
C:Genetics:  
A:Gene: FlyBase:K1p67A  
A:Cross-references: FlyBase:FBgn0004379  
C:Superfamily: Unassigned kinesin-related proteins; kinesin motor domain homology  
C:Keywords: ATP; P-loop  
F:1-118/Domain: kinesin motor domain homology (fragment) <XNOT>

Query Match 100.0%; Score 32; DB 2; Length 118;  
Best Local Similarity 100.0%; Pred. No. 3.1;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATINKSL 7  
DB 52 ATINKSL 58

RESULT 2

S62002  
hypothetical protein YPL120W - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein LPH7W  
C:Species: Saccharomyces cerevisiae  
C>Date: 10-Apr-1996 #sequence\_revision 19-Apr-1996 #text\_change 06-Feb-1998  
C:Accession: S62002  
R:Schlenstedt, G.; Silver, P.A.  
Submitted to the EMBL Data Library, December 1995  
A:Reference number: S61996  
A:Accession: S62002  
A:Molecule type: DNA  
A:Residues: 1-557 <SCH>  
A:Cross-references: EMBL:U43503; NID:g1163087; PID:g1163094; MIPS:YPL120W  
C:Genetics:  
A:Gene: SGD:VPS30  
A:Cross-references: SGD:S0006041; MIPS:YPL120W  
A:Map position: 16L

Query Match 90.6%; Score 29; DB 2; Length 557;  
Best Local Similarity 85.7%; Pred. No. 78;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;



QY 1 ATINKSL 7  
|||||:  
Db 368 ATINKSL 374

## RESULT 3

T41235  
Probable myosin heavy chain - fission yeast (Schizosaccharomyces pombe)  
C;Species: Schizosaccharomyces pombe  
C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 08-Sep-2000  
C;Accession: T41235  
R;Lyne, M.; Harris, D.E.; Murphey, L.D.; Rajandream, M.A.; Barrell, B.G.  
submitted to the EMBL Data Library, January 1999  
A;Reference number: Z21979  
A;Accession: T41235  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-1516 <LYN>  
A;Cross-references: EMBL:AL035075; PIDN:CAA22641.1; GSPDB:GN00060; SPDB:SPCC1919.10C  
A;Experimental source: strain 972h-; cosmid c1919  
C;Genetics:  
A;Gene: SPDB:SPCC1919.10C  
A;Map position: 3  
A;Introns: 10/2; 46/3  
C;Superfamily: myosin MYO2; myosin motor domain homology  
F;76-754/Domain: myosin motor domain homology <MMO>

Query Match 90.6%; Score 29; DB 2; Length 1516;  
Best Local Similarity 85.7%; Pred. No. 2.2e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATINKSL 7  
|||||:  
Db 432 ATINKAL 438

## RESULT 4

QOBEHG  
early nuclear antigen P41 - human herpesvirus 6 (strain UI102 GS)  
C;Species: human herpesvirus 6  
C;Date: 30-Jun-1992 #sequence\_revision 21-Jul-2000 #text\_change 21-Jul-2000  
C;Accession: JQ2007; A39923  
R;Agulnick, A.D.; Thompson, J.R.; Iyengar, S.; Pearson, G.; Ablashi, D.; Ricciardi, R.P.  
J. Gen. Virol. 74, 1003-1009, 1993  
A;Title: Identification of a DNA-binding protein of human herpesvirus 6, a putative DNA  
A;Reference number: JQ2007; MUID:93286553; PMID:8389796  
A;Accession: JQ2007  
A;Molecule type: mRNA  
A;Residues: 1-368 <AGU>  
A;Cross-references: GB:U12003; NID:G325490; PIDN:AAA43855.1; PID:G325491  
A;Experimental source: strain UI102 GS  
R;Chang, C.K.; Balachandran, N.  
J. Virol. 65, 2884-2894, 1991  
A;Title: Identification, characterization, and sequence analysis of a cDNA encoding a ph  
A;Reference number: A39923; MUID:91237802; PMID:1851860  
A;Accession: A39923  
A;Molecule type: DNA  
A;Residues: 1-308, G', 310-320, 'E', 322-361, 'E'AMRRLTDSFILGLAKGAVIFGLYTFRTMGSRPLGQIGVLIV  
SGPLGQKRPILKSTSFREVDKLVKFTVEFLLLFTILWIKMPLSNHLFFLTVVNNVFFKLVNVPKALRAKRWRSYRDL'  
A;Cross-references: GB:M62700  
A;Experimental source: strain UI102 GS  
C;Note: the sequence is revised in GenBank entry HH6P41A, release 117, (PIDN:AAA43853.1)  
C;Superfamily: human herpesvirus 6 P41 protein  
C;Keywords: glycoprotein; phosphoprotein; transmembrane protein  
F;38,51,231,331/binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 87.5%; Score 28; DB 1; Length 368;  
Best Local Similarity 100.0%; Pred. No. 89;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TINKSL 7  
|||||

Db 36 TINKSL 41

## RESULT 5

T09322  
DNA polymerase processivity factor - human herpesvirus 6 (strain UI102)  
C;Species: human herpesvirus 6  
C;Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C;Accession: T09322  
R;Nicholas, J.; Martin, M.  
J. Virol. 68, 597-610, 1994  
A;Title: Nucleotide sequence analysis of a 38.5-kilobase-pair region of the genome of h  
A;Reference number: Z16644; MUID:94118404; PMID:8289364  
A;Accession: T09322  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-403 <NIC>  
A;Cross-references: EMBL:L25528; NID:9451932; PIDN:AAA16735.1; PID:G451953  
C;Genetics:  
A;Gene: BPLF1  
C;Superfamily: human herpesvirus 6 P41 protein

Query Match 87.5%; Score 28; DB 2; Length 403;  
Best Local Similarity 100.0%; Pred. No. 97;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TINKSL 7  
|||||  
Db 71 TINKSL 76

## RESULT 6

T49189  
kinesin heavy chain-like protein - Arabidopsis thaliana  
N;Alternate names: protein MAA21.110  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 02-Jun-2000  
C;Accession: T49189  
R;Rieger, M.; Mueller-Auer, S.; Zipp, M.; Schaefer, M.; Mewes, H.W.; Rudd, S.; Lemcke,  
submitted to the Protein Sequence Database, April 2000  
A;Reference number: Z25018  
A;Accession: T49189  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-439 <RIE>  
A;Cross-references: EMBL:AL163818; GSPDB:GN00061; ATSP:MAA21.110  
A;Experimental source: cultivar Columbia; BAC clone MAA21  
C;Genetics:  
A;Gene: ATSP:MAA21.110  
A;Map position: 3  
A;Introns: 39/3; 74/1; 97/3; 138/3; 147/2; 193/3; 200/2; 273/3; 304/2; 341/3; 401/3

Query Match 87.5%; Score 28; DB 2; Length 439;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TINKSL 7  
|||||  
Db 232 TINKSL 237

## RESULT 7

S66673  
disulfide isomerase-related protein - human  
C;Species: Homo sapiens (man)  
C;Date: 28-Oct-1996 #sequence\_revision 13-Mar-1997 #text\_change 21-Jul-2000  
C;Accession: S66673  
R;Hayano, T.; Kikuchi, M.  
FEBS Lett. 372, 210-214, 1995  
A;Title: Molecular cloning of the cDNA encoding a novel protein disulfide isomerase-re  
A;Reference number: S66673; MUID:96000209; PMID:7556671  
A;Accession: S66673  
A;Status: preliminary

Query Match 87.5%; Score 28; DB 2; Length 636;  
Best Local Similarity 71.4%; Pred. No. 1.9e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATINKSL 7  
DB 332 ATINQSL 338  
|||:|

RESULT 10  
AE1240  
translation initiation factor IF-2 homolog infB [imported] - Listeria monocytogenes (strain  
C;Species: Listeria monocytogenes  
C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001  
C;Accession: AE1240  
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker  
; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, F.  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; N  
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlant  
A;Title: Comparative genomics of Listeria species.  
A;Reference number: AB1077; MUID:21537279; PMID:11679669  
A;Accession: AE1240  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-779 <GLA>  
A;Cross-references: GB:NC\_003210; PID:G16410741; GSPDB:GN00177  
A;Experimental source: strain EGD-e  
C;Genetics:  
A;Gene: infB  
C;Superfamily: translation initiation factor IF-2; translation elongation factor Tu hom

Query Match 87.5%; Score 28; DB 2; Length 779;  
Best Local Similarity 85.7%; Pred. No. 1.9e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATINKSL 7  
DB 228 ATINQSL 234  
|||:|

RESULT 11  
All602  
translation initiation factor IF-2 homolog infB [imported] - Listeria innocua (strain  
C;Species: Listeria innocua  
C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001  
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker  
; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, F.  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; N  
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlant  
A;Title: Comparative genomics of Listeria species.  
A;Reference number: AB1077; MUID:21537279; PMID:11679669  
A;Accession: All602  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-782 <GLA>  
A;Cross-references: GB:AL592022; PID:NCAC96593.1; PID:g16413835; GSPDB:GN00178  
A;Experimental source: strain Clip11262  
C;Genetics:  
A;Gene: infB  
C;Superfamily: translation initiation factor IF-2; translation elongation factor Tu hom

Query Match 87.5%; Score 28; DB 2; Length 782;  
Best Local Similarity 85.7%; Pred. No. 1.9e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATINKSL 7  
DB 228 ATINQSL 234  
|||:|

## RESULT 12

S08243  
 beta-glucosidase (EC 3.2.1.21) - Ruminococcus albus  
 C:Species: Ruminococcus albus  
 C:Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 15-Oct-1999  
 C:Accession: S08243  
 R:Ohnishi, K.; Takano, M.; Shimizu, S.  
 C:Nucleic Acids Res. 18, 671, 1990  
 A:Title: DNA sequence of a beta-glucosidase from Ruminococcus albus.  
 A:Reference number: S08243; MUID:90175009; PMID:2106673  
 A:Molecule type: DNA  
 A:Accession: S08243  
 A:Residues: 1-947 <OHM>  
 A:Cross-references: EMBL:X15415; NID:G45967; PIDN:CAA33461.1; PID:G45968  
 C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 87.5%; Score 28; DB 2; Length 947;  
 Best Local Similarity 85.7%; Pred. No. 2.3e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATINKSL 7  
 |||||  
 Db 543 ATFNKSL 549

## RESULT 13

S59393  
 probable membrane protein YLR247c - yeast (Saccharomyces cerevisiae)  
 N:Alternate names: hypothetical protein L9672.14  
 C:Species: Saccharomyces cerevisiae  
 C:Date: 30-Nov-1995 #sequence\_revision 23-Feb-1996 #text\_change 19-Apr-2002  
 C:Accession: S59393  
 R:Johnson, D.  
 submitted to the EMBL Data Library, February 1995  
 A:Description: The sequence of S. cerevisiae cosmid 9672.  
 A:Reference number: S59386  
 A:Accession: S59393  
 A:Molecule type: DNA  
 A:Residues: 1-1556 <JOH>  
 A:Cross-references: EMBL:U20865; NID:G662330; PID:G662344; GSPDB:GN00012; MIPS:YLR247C  
 A:Experimental source: strain S288C (AB972)  
 C:Genetics:

Query Match 87.5%; Score 28; DB 2; Length 1556;  
 Best Local Similarity 100.0%; Pred. No. 3.8e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TINKSL 7  
 |||||  
 Db 1198 TINKSL 1203

## RESULT 14

T38660  
 probable transcription regulator protein - fission yeast (Schizosaccharomyces pombe)  
 C:Species: Schizosaccharomyces pombe  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
 C:Accession: T38660  
 R:Murphy, L.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.  
 submitted to the EMBL Data Library, August 1997  
 A:Reference number: Z21804  
 A:Accession: T38660  
 A:Status: preliminary; translated from GB/EMBL/DDBJ  
 A:Molecule type: DNA

A:Residues: 1-1588 <MUR>  
 A:Cross-references: EMBL:AL109739; NID:el534774; PIDN:CAB52274.1; GSPDB:GN00066; SPDB:S  
 A:Experimental source: strain 972h-; cosmid c343  
 C:Genetics:  
 A:Gene: SPDB:SPAC343.11c  
 A:Map position: 1

Query Match 87.5%; Score 28; DB 2; Length 1588;  
 Best Local Similarity 100.0%; Pred. No. 3.9e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATINKS 6  
 |||||  
 Db 175 ATINKS 180

## RESULT 15

AF2140  
 hypothetical protein all2677 [imported] - Nostoc sp. (strain PCC 7120)  
 C:Species: Nostoc sp. PCC 7120  
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
 C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002  
 C:Accession: AF2140  
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguch  
 R:Kakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata,  
 DNA Res. 8, 205-213, 2001  
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ar  
 A:Reference number: AB1807; MUID:21595285; PMID:11759840  
 A:Accession: AF2140  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-225 <KUR>  
 A:Cross-references: GB:BA000019; PIDN:BA074376.1; PID:gl7131770; GSPDB:GN00179  
 A:Experimental source: strain PCC 7120  
 C:Genetics:

A:Gene: all2677

Query Match 84.4%; Score 27; DB 2; Length 225;  
 Best Local Similarity 71.4%; Pred. No. 93;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATINKSL 7  
 |||||  
 Db 215 ATINKSL 221

Search completed: February 11, 2004, 17:11:34  
 Job time : 11.3333 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 17:09:56 ; Search time 24.6667 Seconds  
(without alignments)  
59.419 Million cell updates/sec

Title: US-09-901-187C-3

Perfect score: 32

Sequence: 1 ATINKSL 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 801455 seqs, 209382283 residues

Total number of hits satisfying chosen parameters: 801455

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:  
1: /cgn2\_6/prodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/prodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/prodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/prodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/prodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/prodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/prodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/prodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/prodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/prodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/prodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/prodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/prodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/prodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/prodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/prodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/prodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/prodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	32	100.0	7	10	US-09-901-187B-3
2	29	90.6	557	12	Sequence 3, Appli
3	28	90.6	1537	12	Sequence 22785, A
4	28	87.5	149	9	Sequence 22785, A
5	28	87.5	204	12	Sequence 40466, A
6	28	87.5	234	12	Sequence 1816, Ap
7	28	87.5	464	12	Sequence 4130, Ap
8	28	87.5	521	12	Sequence 119, App
9	28	87.5	526	12	Sequence 48, Appl
10	28	87.5	553	12	Sequence 32, Appl
11	28	87.5	553	12	Sequence 5944, Ap
12	28	87.5	554	11	Sequence 370, App
13	28	87.5	1556	12	Sequence 6, Appli
14	27	84.4	165	9	Sequence 1796, Ap
15	27	84.4	431	12	Sequence 45486, A
					Sequence 3314, Ap

16	27	84.4	491	12	US-10-369-493-6081
17	27	84.4	506	10	Sequence 90, Appl
18	27	84.4	506	12	Sequence 1559, Ap
19	27	84.4	511	12	Sequence 46, Appl
20	27	84.4	536	12	Sequence 14, Appl
21	27	84.4	574	15	Sequence 11937, A
22	27	84.4	873	12	Sequence 2449, Ap
23	27	84.4	1103	10	Sequence 1, Appli
24	27	84.4	1103	12	Sequence 1, Appli
25	26	81.2	46	9	Sequence 36397, A
26	26	81.2	76	15	Sequence 32, Appl
27	26	81.2	79	9	Sequence 68, Appl
28	26	81.2	91	12	Sequence 22, Appl
29	26	81.2	91	12	Sequence 34, Appl
30	26	81.2	122	7	Sequence 28, Appl
31	26	81.2	177	9	Sequence 7, Appli
32	26	81.2	185	12	Sequence 172, App
33	26	81.2	235	12	Sequence 5, Appli
34	26	81.2	261	12	Sequence 25, Appl
35	26	81.2	266	12	Sequence 20, Appl
36	26	81.2	271	12	Sequence 18215, A
37	26	81.2	277	12	Sequence 8, Appli
38	26	81.2	293	12	Sequence 140, App
39	26	81.2	303	12	Sequence 4, Appli
40	26	81.2	319	12	Sequence 23683, A
41	26	81.2	335	12	Sequence 16, Appl
42	26	81.2	410	10	Sequence 1074, Ap
43	26	81.2	433	11	Sequence 4, Appli
44	26	81.2	464	12	Sequence 18, Appl
45	26	81.2	476	10	Sequence 27, Appl

ALIGNMENTS

RESULT 1  
US-09-901-187B-3  
; Sequence 3, Application US/09901187B  
; Patent No. US2002015144A1  
; GENERAL INFORMATION:  
; APPLICANT: Panacea Pharmaceuticals, Inc.  
; APPLICANT: Wolozin, Benjamin  
; APPLICANT: Ostretova-Golts, Natalie  
; APPLICANT: Lebowitz, Michael S  
; TITLE OF INVENTION: Methods for Preventing Neural Tissue Damage and for the Treatment  
; FILE REFERENCE: PAN01/002US  
; CURRENT APPLICATION NUMBER: US/09/901.187B  
; CURRENT FILING DATE: 2001-07-09  
; PRIOR APPLICATION NUMBER: US 60/217,319  
; PRIOR FILING DATE: 2000-07-07  
; PRIOR APPLICATION NUMBER: US 60/279,199  
; PRIOR FILING DATE: 2001-03-28  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 3  
; LENGTH: 7  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-901-187B-3

Query Match 100.0%; Score 32; DB 10; Length 7;  
Best Local Similarity 100.0%; Pred. No. 7e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATINKSL 7  
DB 1 ATINKSL 7

RESULT 2  
US-10-369-493-22479  
; Sequence 22479, Application US/10369493

```

; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 22479
; LENGTH: 557
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-369-493-22479

Query Match          90.6%; Score 29; DB 12; Length 557;
Best Local Similarity 85.7%; Pred. No. 2.1e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 ATINKSL 7
DB      368 ATINKSL 374

RESULT 3
US-10-369-493-22785
; Sequence 22785, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 22785
; LENGTH: 1537
; TYPE: PRT
; ORGANISM: Schizosaccharomyces pombe
; NAME/KEY: unsure
; LOCATION: (1)..(1537)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-22785

Query Match          90.6%; Score 29; DB 12; Length 1537;
Best Local Similarity 85.7%; Pred. No. 6.4e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 ATINKSL 7
DB      453 ATINKSL 459

RESULT 4
US-09-864-761-40466
; Sequence 40466, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.

```

```

; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wenheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 40466
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC024196.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.92
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1
; OTHER INFORMATION: EST HUMAN HIT: AU121127.1, EVALUE 3.00e-21
; OTHER INFORMATION: SWISSPROT HIT: P19013, EVALUE 5.00e-21
US-09-864-761-40466

Query Match          87.5%; Score 28; DB 9; Length 149;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 TINKSL 7
DB      108 TINKSL 113

RESULT 5

```

```

US-10-094-749-1816
; Sequence 1816, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; PRIOR FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1816
; LENGTH: 204
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-1816

Query Match      87.5%; Score 28; DB 12; Length 204;
Best Local Similarity 71.4%; Pred. No. 1.2e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 ATINKSL 7
      |||||
Db      27 ATVNKL 33

RESULT 6
US-10-264-049-4130
; Sequence 4130, Application US/10264049
; Publication No. US20040005579A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA133PI
; CURRENT APPLICATION NUMBER: US/10/264,049
; PRIOR FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/18569
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: US 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 4360
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 4130
; LENGTH: 234
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (74)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC FEATURE

```

```

; LOCATION: (177)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (199)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (209)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (213)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
US-10-264-049-4130

Query Match      87.5%; Score 28; DB 12; Length 234;
Best Local Similarity 71.4%; Pred. No. 1.4e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 ATINKSL 7
      |||||
Db      60 ATVNKL 66

RESULT 7
US-10-161-051-119
; Sequence 119, Application US/10161051
; Publication No. US20030152945A1
; GENERAL INFORMATION:
; APPLICANT: Peter Deak
; APPLICANT: David M. Glover
; APPLICANT: Carol Midgley
; TITLE OF INVENTION: Cell cycle progression proteins
; FILE REFERENCE: CCI-021CP
; CURRENT APPLICATION NUMBER: US/10/161,051
; CURRENT FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: GB 0007268.6
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 119
; LENGTH: 464
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-10-161-051-119

Query Match      87.5%; Score 28; DB 12; Length 464;
Best Local Similarity 85.7%; Pred. No. 3e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 ATINKSL 7
      |||||
Db      273 ASINKSL 279

RESULT 8
US-10-015-115-48
; Sequence 48, Application US/10015115
; Publication No. US20030207800A1
; GENERAL INFORMATION:
; APPLICANT: Malyankar, Uriel M
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Zerhusen, Bryan D
; APPLICANT: Patturajan, Meera
; APPLICANT: Guo, Xiaojia
; APPLICANT: Kekuda, Ramesha
; APPLICANT: Gangolli, Esha A
; APPLICANT: Shinkets, Richard A
; APPLICANT: Taupier, Raymond J
; APPLICANT: Li, Li
; APPLICANT: Padigaru, Muralidhara

```

```

; TITLE OF INVENTION: Proteins, Polynucleotides Encoding Them and Methods of
; TITLE OF INVENTION: Using the Same
; FILE REFERENCE: 21402-211
; CURRENT APPLICATION NUMBER: US/10/015,115
; CURRENT FILING DATE: 2002-09-23
; PRIOR APPLICATION NUMBER: 60/248,153
; PRIOR FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: 60/249,598
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/264,240
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 60/266,127
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/269,562
; PRIOR FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: 60/304,348
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/309,261
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/313,283
; PRIOR FILING DATE: 2001-08-17
; NUMBER OF SEQ ID NOS: 205
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 48
; LENGTH: 521
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-015-115-48

Query Match      87.5%; Score 28; DB 12; Length 521;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 TINKSL 7
DB      108 TINKSL 113

RESULT 9
US-10-231-913-32
; Sequence 32, Application US/10231913
; Publication No. US20040005576A1
; GENERAL INFORMATION:
; APPLICANT: Guo, Xiaojia S.
; APPLICANT: Li, Li
; APPLICANT: Patturajan, Meera
; APPLICANT: Shimkets, Richard A.
; APPLICANT: Casman, Stacie J.
; APPLICANT: Malyankar, Uriel M.
; APPLICANT: Tchernev, Velizar T.
; APPLICANT: Vernet, Corine A.
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Shency, Suresh G.
; APPLICANT: Alsobrook II, John P.
; APPLICANT: Edinger, Schlomit
; APPLICANT: Peyman, John A.
; APPLICANT: Stone, David J.
; APPLICANT: Ellerman, Karen
; APPLICANT: Gangolli, Esha A.
; APPLICANT: Boldog, Ference L.
; APPLICANT: Colman, Steven D.
; APPLICANT: Eisen, Andrew J.
; APPLICANT: Liu, Xiaohong
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Spaderma, Steven K.
; APPLICANT: Zerhusen, Bryan D.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-216
; CURRENT APPLICATION NUMBER: US/10/231,913
; CURRENT FILING DATE: 2002-08-30
; PRIOR APPLICATION NUMBER: 60/251,660
; PRIOR FILING DATE: 2000-12-06
; PRIOR APPLICATION NUMBER: 60/255,029

```

```

; PRIOR FILING DATE: 2000-12-12
; PRIOR APPLICATION NUMBER: 60/260,326
; PRIOR FILING DATE: 2001-01-08
; PRIOR APPLICATION NUMBER: 60/263,800
; PRIOR FILING DATE: 2001-01-24
; PRIOR APPLICATION NUMBER: 60/269,942
; PRIOR FILING DATE: 2001-02-20
; PRIOR APPLICATION NUMBER: 60/286,183
; PRIOR FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: 60/313,627
; PRIOR FILING DATE: 2001-08-20
; PRIOR APPLICATION NUMBER: 60/319,712
; PRIOR FILING DATE: 2001-09-12
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 32
; LENGTH: 526
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-231-913-32

Query Match      87.5%; Score 28; DB 12; Length 526;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 TINKSL 7
DB      108 TINKSL 113

RESULT 10
US-10-369-493-5944
; Sequence 9944, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xiaofeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 5944
; LENGTH: 550
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-369-493-5944

Query Match      87.5%; Score 28; DB 12; Length 550;
Best Local Similarity 71.4%; Pred. No. 3.6e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 ATINKSL 7
DB      400 STINKSL 406

RESULT 11
US-10-295-027-370
; Sequence 370, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard

```

APPLICANT: Hevezi, Peter A.  
APPLICANT: Mack, David H.  
APPLICANT: Murray, Richard  
APPLICANT: Watson, Susan R.  
APPLICANT: Eos Biotechnology, Inc.  
TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and  
TITLE OF INVENTION: Methods of Screening for Modulators of Cancer  
FILE REFERENCE: 018501-012500US  
CURRENT APPLICATION NUMBER: US/10/295,027  
CURRENT FILING DATE: 2002-11-13  
PRIOR APPLICATION NUMBER: US 09/663,733  
PRIOR FILING DATE: 2000-09-15  
PRIOR APPLICATION NUMBER: US 60/350,666  
PRIOR FILING DATE: 2001-11-13  
PRIOR APPLICATION NUMBER: US 60/335,394  
PRIOR FILING DATE: 2001-11-15  
PRIOR APPLICATION NUMBER: US 60/332,464  
PRIOR FILING DATE: 2001-11-21  
PRIOR APPLICATION NUMBER: US 60/334,393  
PRIOR FILING DATE: 2001-11-29  
PRIOR APPLICATION NUMBER: US 60/340,376  
PRIOR FILING DATE: 2001-12-14  
PRIOR APPLICATION NUMBER: US 60/347,211  
PRIOR FILING DATE: 2002-01-08  
PRIOR APPLICATION NUMBER: US 60/347,349  
PRIOR FILING DATE: 2002-01-10  
PRIOR APPLICATION NUMBER: US 60/355,250  
PRIOR FILING DATE: 2002-02-08  
PRIOR APPLICATION NUMBER: US 60/356,714  
PRIOR FILING DATE: 2002-02-13  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 1386  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 370  
LENGTH: 553  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-295-027-370

Query Match 87.5%; Score 28; DB 12; Length 553;  
Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATINKS 6  
DB 459 ATINKS 464

RESULT 12  
US-09-902-939-6  
Sequence 6, Application US/09902939  
Publication No. US20030087850A1  
GENERAL INFORMATION:  
APPLICANT: Phillip DeHazy  
APPLICANT: William Chen  
TITLE OF INVENTION: GENE THERAPY FOR DRY EYE SYNDROME  
FILE REFERENCE: 2055/OH020-US0  
CURRENT APPLICATION NUMBER: US/09/902,939  
CURRENT FILING DATE: 2001-07-10  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 6  
LENGTH: 554  
TYPE: PRT  
ORGANISM: mouse  
US-09-902-939-6

Query Match 87.5%; Score 28; DB 11; Length 554;  
Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATINKS 6  
DB 459 ATINKS 464

Db 460 ATINKS 465

RESULT 13  
US-10-369-493-1796  
Sequence 1796, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10/52052B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 1796  
LENGTH: 1556  
TYPE: PRT  
ORGANISM: Saccharomyces cerevisiae  
US-10-369-493-1796

Query Match 87.5%; Score 28; DB 12; Length 1556;  
Best Local Similarity 100.0%; Pred. No. 1.1e+03;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TINKSL 7  
DB 1198 TINKSL 1203

RESULT 14  
US-09-864-761-45486  
Sequence 45486, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharron G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aecmics-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663



Job time : 26.6667 secs

;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00662  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00661  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00670  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: US 60/234,687  
;; PRIOR FILING DATE: 2000-09-21  
;; PRIOR APPLICATION NUMBER: US 09/608,408  
;; PRIOR FILING DATE: 2000-06-30  
;; PRIOR APPLICATION NUMBER: US 09/774,203  
;; PRIOR FILING DATE: 2001-01-29  
;; NUMBER OF SEQ ID NOS: 49117  
;; SOFTWARE: Annotmax Sequence Listing Engine vers. 1.1  
;; SEQ ID NO 45486  
;; LENGTH: 165  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; OTHER INFORMATION: MAP TO AC024196.2  
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.45  
;; OTHER INFORMATION: SWISSPROT HIT: P48669, EVALUE 4.00e-19  
;; OTHER INFORMATION: EST\_HUMAN HIT: AW262317.1, EVALUE 2.00e-17  
US-09-864-761-45486

Query Match 84.4%; Score 27; DB 9; Length 165;  
Best Local Similarity 83.3%; Pred. No. 1.6e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TINKSL 7  
|:||||  
Db 124 TVNKSL 129

RESULT 15  
US-10-369-493-3314  
; Sequence 3314, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 3314  
; LENGTH: 431  
; TYPE: PRT  
; ORGANISM: Neurospora crassa  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)...(431)  
; OTHER INFORMATION: unsure at all Xaa locations  
US-10-369-493-3314

Query Match 84.4%; Score 27; DB 12; Length 431;  
Best Local Similarity 85.7%; Pred. No. 4.7e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATINKSL 7  
|:||||  
Db 283 ANINKSL 289

Search completed: February 11, 2004, 17:54:06

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 16:56:09 ; Search time 10.4167 Seconds  
(without alignments)  
28.433 Million cell updates/sec

Title: US-09-901-187C-3

Perfect score: 32

Sequence: 1 ATINKSL 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	27	84.4	114	4	US-09-252-991A-21141
2	27	84.4	234	4	US-09-328-352-6763
3	27	84.4	357	4	US-09-718-841-4
4	27	84.4	357	4	US-09-718-810-4
5	27	84.4	425	4	US-09-328-352-7263
6	27	84.4	563	4	US-09-718-841-2
7	27	84.4	563	4	US-09-718-810-2
8	27	84.4	1103	3	US-09-162-373-1
9	27	84.4	1103	3	US-09-467-946-1
10	27	84.4	1690	4	US-09-595-684B-39
11	26	81.2	79	1	US-08-665-292-68
12	26	81.2	79	3	US-09-291-620-68
13	26	81.2	83	3	US-09-382-155-12
14	26	81.2	83	3	US-09-074-044A-12
15	26	81.2	91	4	US-09-983-502-22
16	26	81.2	91	4	US-09-983-502-34
17	26	81.2	91	4	US-09-516-747-22
18	26	81.2	91	4	US-09-516-747-34
19	26	81.2	91	5	PCT-US96-10521-22
20	26	81.2	91	5	PCT-US96-10521-34
21	26	81.2	180	3	US-09-382-155-18
22	26	81.2	180	3	US-09-074-044A-18
23	26	81.2	220	2	US-08-807-200-2
24	26	81.2	220	3	US-09-001-777-2
25	26	81.2	235	4	US-08-983-502-5
26	26	81.2	235	4	US-09-516-747-5
27	26	81.2	235	5	PCT-US96-10521-5

ALIGNMENTS

RESULT 1  
US-09-252-991A-21141  
; Sequence 21141, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 21141  
; LENGTH: 114  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-21141

Query Match 84.4%; Score 27; DB 4; Length 114;  
Best Local Similarity 83.3%; Pred. No. 39;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 2 TINKSL 7  
Db 1 TVNKSL 6

RESULT 2  
US-09-328-352-6763  
; Sequence 6763, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 6763  
; LENGTH: 234  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-6763

Query Match 84.4%; Score 27; DB 4; Length 234;

Best Local Similarity 83.3%; Pred. No. 90;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATINKS 6  
Db 219 ATINKS 224

## RESULT 3

US-09-718-841-4  
; Sequence 4, Application US/09718841  
; Patent No. 633184  
; GENERAL INFORMATION:  
; APPLICANT: Beraud, Christophe  
; APPLICANT: Freedman, Richard  
; TITLE OF INVENTION: NO. 633184e1 motor proteins and methods for  
; TITLE OF INVENTION: their use  
; FILE REFERENCE: 1051  
; CURRENT APPLICATION NUMBER: US/09/718,841  
; CURRENT FILING DATE: 2000-11-22  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 357  
; TYPE: PRT  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (16)...(37)  
; OTHER INFORMATION: Xaa = any amino acid  
US-09-718-841-4

Query Match 84.4%; Score 27; DB 4; Length 357;  
Best Local Similarity 85.7%; Pred. No. 1.5e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATINKSL 7  
Db 271 ANINKSL 277

## RESULT 4

US-09-718-810-4  
; Sequence 4, Application US/09718810  
; Patent No. 6420162  
; GENERAL INFORMATION:  
; APPLICANT: Beraud, Christophe  
; APPLICANT: Freedman, Richard  
; TITLE OF INVENTION: NO. 6420162e1 motor proteins and methods for  
; TITLE OF INVENTION: their use  
; FILE REFERENCE: 1051  
; CURRENT APPLICATION NUMBER: US/09/718,810  
; CURRENT FILING DATE: 2000-11-22  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 357  
; TYPE: PRT  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (16)...(37)  
; OTHER INFORMATION: Xaa = any amino acid  
US-09-718-810-4

Query Match 84.4%; Score 27; DB 4; Length 357;  
Best Local Similarity 85.7%; Pred. No. 1.5e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATINKSL 7  
Db 271 ANINKSL 277

## RESULT 5

US-09-328-352-7263  
; Sequence 7263, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gaty L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 7263  
; LENGTH: 425  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-7263

Query Match 84.4%; Score 27; DB 4; Length 425;  
Best Local Similarity 83.3%; Pred. No. 1.8e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 TINKSL 7  
Db 361 TVNKSL 366

## RESULT 6

US-09-718-841-2  
; Sequence 2, Application US/09718841  
; Patent No. 633184  
; GENERAL INFORMATION:  
; APPLICANT: Beraud, Christophe  
; APPLICANT: Freedman, Richard  
; TITLE OF INVENTION: NO. 633184e1 motor proteins and methods for  
; TITLE OF INVENTION: their use  
; FILE REFERENCE: 1051  
; CURRENT APPLICATION NUMBER: US/09/718,841  
; CURRENT FILING DATE: 2000-11-22  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 563  
; TYPE: PRT  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (16)...(37)  
; OTHER INFORMATION: Xaa = any amino acid  
; NAME/KEY: VARIANT  
; LOCATION: (391)...(412)  
; OTHER INFORMATION: Xaa = any amino acid  
US-09-718-841-2

Query Match 84.4%; Score 27; DB 4; Length 563;  
Best Local Similarity 85.7%; Pred. No. 2.5e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATINKSL 7  
Db 271 ANINKSL 277

## RESULT 7

US-09-718-810-2  
; Sequence 2, Application US/09718810  
; Patent No. 6420162  
; GENERAL INFORMATION:  
; APPLICANT: Beraud, Christophe  
; APPLICANT: Freedman, Richard  
; TITLE OF INVENTION: NO. 6420162e1 motor proteins and methods for  
; TITLE OF INVENTION: their use  
US-09-718-810-2

```

; FILE REFERENCE: 1051
; CURRENT APPLICATION NUMBER: US/09/718,810
; CURRENT FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 563
; TYPE: PRT
; ORGANISM: Human
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (16)...(37)
; OTHER INFORMATION: Xaa = any amino acid
; NAME/KEY: VARIANT
; LOCATION: (391)...(412)
; OTHER INFORMATION: Xaa = any amino acid
US-09-718-810-2

Query Match      84.4%; Score 27; DB 4; Length 563;
Best Local Similarity 85.7%; Pred. No. 2.5e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 ATINKSL 7
DB      271 ANINKSL 277

RESULT 8
US-09-162-373-1
; Sequence 1, Application US/09162373
; Patent No. 6013454
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Patterson, Chandra
; TITLE OF INVENTION: KINESIN-LIKE MOTOR PROTEIN
; FILE REFERENCE: PF-0593 US
; CURRENT APPLICATION NUMBER: US/09/162,373
; CURRENT FILING DATE: 1998-09-28
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PERL Program
; SEQ ID NO 1
; LENGTH: 1103
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: 1281811
US-09-162-373-1

Query Match      84.4%; Score 27; DB 3; Length 1103;
Best Local Similarity 85.7%; Pred. No. 5.5e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 ATINKSL 7
DB      269 ANINKSL 275

RESULT 9
US-09-467-946-1
; Sequence 1, Application US/09467946
; Patent No. 6245594
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Patterson, Chandra
; TITLE OF INVENTION: KINESIN-LIKE MOTOR PROTEIN
; FILE REFERENCE: PF-0593 US
; CURRENT APPLICATION NUMBER: US/09/467,946
; CURRENT FILING DATE: 1999-12-21
; EARLIER APPLICATION NUMBER: 09/162,373
; EARLIER FILING DATE: 1998-09-28
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PERL Program
; SEQ ID NO 1
; LENGTH: 1103
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: 1281811
US-09-467-946-1

Query Match      84.4%; Score 27; DB 3; Length 1103;
Best Local Similarity 85.7%; Pred. No. 5.5e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 ATINKSL 7
DB      269 ANINKSL 275

RESULT 10
US-09-595-684B-39
; Sequence 39, Application US/09595684B
; Patent No. 6544766
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Ohashi, Cara
; APPLICANT: Sakowicz, Roman
; APPLICANT: Vaisberg, Eugeni
; APPLICANT: Wood, Kenneth
; APPLICANT: Yu, Ming
; TITLE OF INVENTION: Human kinesins and methods of producing
; FILE REFERENCE: cytop036
; CURRENT APPLICATION NUMBER: US/09/595,684B
; CURRENT FILING DATE: 2002-06-24
; PRIOR APPLICATION NUMBER: 09/295,612
; PRIOR FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 39
; LENGTH: 1690
; TYPE: PRT
; ORGANISM: Human
; FEATURE:
; OTHER INFORMATION: 1281811
US-09-595-684B-39

Query Match      84.4%; Score 27; DB 4; Length 1690;
Best Local Similarity 85.7%; Pred. No. 9.2e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 ATINKSL 7
DB      269 ANINKSL 275

RESULT 11
US-08-665-220-68
; Sequence 68, Application US/08665220
; Patent No. 5786173
; GENERAL INFORMATION:
; APPLICANT: Alnemri, Emad S.
; APPLICANT: Fernandes-Alnemri, Teresa
; APPLICANT: Litwack, Gerald
; APPLICANT: Armstrong, Robert
; APPLICANT: Tomaselli, Kevin
; TITLE OF INVENTION: Mch4 and Mch5, Apoptotic Proteases,
; TITLE OF INVENTION: Nucleic Acids Encoding and Methods of Use
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California

```

```

; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/665,220
; APPLICATION NUMBER: US/08/665,220
; FILING DATE: 14-JUN-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/618,408
; FILING DATE: 19-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-ID 2165
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 79 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1..79
; OTHER INFORMATION: /note= "Mch5 B"
; US-08-665-220-68
;
; Query Match 81.2%; Score 26; DB 1; Length 79;
; Best Local Similarity 85.7%; Pred. No. 43;
; Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
;
; QY 1 ATINKSL 7
; Db 70 AQINKSL 76
;
; RESULT 12
; US-09-291-692-68
; Sequence 68, Application US/09291692
; Patent No. 6287795
; GENERAL INFORMATION:
; APPLICANT: Alnemri, Emad S.
; APPLICANT: Fernandes-Alnemri, Teresa
; APPLICANT: Litwack, Gerald
; APPLICANT: Armstrong, Robert
; APPLICANT: Tomaselli, Kevin
; TITLE OF INVENTION: MCH4 AND MCH5, APOPTOTIC PROTEASE,
; NUCLEIC ACIDS ENCODING AND METHODS OF USE
; NUMBER OF SEQUENCES: 75
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/291,692
; FILING DATE: 04-13-1999
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Christiansen, William T.
; REGISTRATION NUMBER: 44,614
;
; COUNTRY: United States
; ZIP: 92122
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-8031
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 79 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1..79
; OTHER INFORMATION: /note= "Mch5 B"
; US-09-291-692-68
;
; Query Match 81.2%; Score 26; DB 3; Length 79;
; Best Local Similarity 85.7%; Pred. No. 43;
; Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
;
; QY 1 ATINKSL 7
; Db 70 AQINKSL 76
;
; RESULT 13
; US-09-382-155-12
; Sequence 12, Application US/09382155B
; Patent No. 6160095
; GENERAL INFORMATION:
; APPLICANT: CHAUDHARY, PREET M
; APPLICANT: HOOD, LEROY
; TITLE OF INVENTION: PROTEINS CAPABLE OF REGULATING NF-KB, JNK AND APOPTOSIS
; FILE REFERENCE: Chaudhary
; CURRENT FILING DATE: 1999-08-24
; EARLIER APPLICATION NUMBER: 09/074,044
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; TYPE: PRT
; LENGTH: 83
; ORGANISM: Homo sapiens
; US-09-382-155-12
;
; Query Match 81.2%; Score 26; DB 3; Length 83;
; Best Local Similarity 85.7%; Pred. No. 46;
; Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
;
; QY 1 ATINKSL 7
; Db 68 AQINKSL 74
;
; RESULT 14
; US-09-074-044A-12
; Sequence 12, Application US/09074044A
; Patent No. 6207458
; GENERAL INFORMATION:
; APPLICANT: CHAUDHARY, PREET M
; APPLICANT: HOOD, LEROY
; TITLE OF INVENTION: PROTEINS CAPABLE OF REGULATING NF-KB, JNK AND
; APOPTOSIS PATHWAYS AND METHODS OF USING THE SAME
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HOVEY, WILLIAMS, TIMMONS & COLLINS
; STREET: 2405 GRAND BLVD., SUITE 400
; CITY: KANSAS CITY
; STATE: MISSOURI
; COUNTRY: USA
; ZIP: 64108

```

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/074,044A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: COLLINS, JOHN M  
REGISTRATION NUMBER: 26,262  
REFERENCE/DOCKET NUMBER: 26588  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 816/474-9050  
TELEFAX: 816/474-9057  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 83 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: not relevant  
MOLECULE TYPE: protein  
FRAGMENT TYPE: internal  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
US-09-074-044A-12

Query Match 81.2%; Score 26; DB 3; Length 83;  
Best Local Similarity 85.7%; Pred. No. 46;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATINKSL 7  
DB 68 AQINKSL 74

RESULT 15  
US-08-983-502-22  
Sequence 22, Application US/08983502  
Patent No. 6399327  
GENERAL INFORMATION:  
APPLICANT: David WALLACH  
APPLICANT: Mark P. BOLDIN  
APPLICANT: Tanya M. GONCHAROV  
APPLICANT: Yuri V. GOLTSEV  
TITLE OF INVENTION: MODULATORS OF THE FUNCTION OF PAS RECEPTORS  
TITLE OF INVENTION: AND OTHER PROTEINS  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Browdy and Neimark  
STREET: 419 Seventh Street N.W., Ste. 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/983,502  
FILING DATE: 16-JAN-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/10521  
FILING DATE: 14-JUN-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: IL 114,615  
FILING DATE: 16-JUL-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: IL 114,986  
FILING DATE: 17-AUG-1995

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: IL 115,319  
FILING DATE: 14-SEP-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: IL 116,588  
FILING DATE: 27-DEC-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: IL 117,932  
FILING DATE: 16-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Browdy, Roger L.  
REGISTRATION NUMBER: 25,618  
REFERENCE/DOCKET NUMBER: WALLACH=19  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 628-5197  
TELEFAX: (202) 737-3528  
INFORMATION FOR SEQ ID NO: 22:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 91 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-983-502-22

Query Match 81.2%; Score 26; DB 4; Length 91;  
Best Local Similarity 85.7%; Pred. No. 51;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATINKSL 7  
DB 36 AQINKSL 42

Search completed: February 11, 2004, 17:13:34  
Job time : 11.4167 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 16:51:33 ; Search time 32.25 Seconds  
(without alignments)  
34.452 Million cell updates/sec

Title: US-09-901-187C-3

Perfect score: 32

Sequence: 1 AINXSL 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- A\_Geneseq\_19Jun03.\*
- 1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*
  - 2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*
  - 3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*
  - 4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*
  - 5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*
  - 6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*
  - 7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*
  - 8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*
  - 9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*
  - 10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*
  - 11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*
  - 12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*
  - 13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*
  - 14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*
  - 15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*
  - 16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*
  - 17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*
  - 18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*
  - 19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*
  - 20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*
  - 21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*
  - 22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*
  - 23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*
  - 24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	32	100.0	7	23	AAE14548
2	30	93.8	74	22	AAU39185
3	28	87.5	139	22	AAO02587
4	28	87.5	149	22	ABG56567
5	28	87.5	149	22	ABB41125
6	28	87.5	149	22	ABB25168
7	28	87.5	149	22	AAU61981
8	28	87.5	149	22	AAU74785
9	28	87.5	149	22	AAU34901

10	28	87.5	149	23	ABG44576	Human peptide enco
11	28	87.5	130	19	AAU50013	Human hyaluronan s
12	28	87.5	199	22	AAU27524	Human G-Protein Co
13	28	87.5	234	23	ABP42998	Human ovarian anti
14	28	87.5	339	22	ABU53124	Intracellular traf
15	28	87.5	395	22	ABG24679	Novel human diagno
16	28	87.5	398	21	AAU21667	Arabidopsis thalia
17	28	87.5	452	21	AAU21666	Arabidopsis thalia
18	28	87.5	464	23	AAU47592	Arabidopsis thalia
19	28	87.5	469	21	AAU21665	Human NOVI0C prote
20	28	87.5	521	23	ABG94649	Protein of NOVA 13
21	28	87.5	526	23	ABU04655	Human cytoskeleton
22	28	87.5	529	23	ABR33482	Human bladder can
23	28	87.5	553	24	ABR48209	Lung cancer-associ
24	28	87.5	553	24	ABU56490	Lung cancer-associ
25	28	87.5	553	24	ABU56686	Murine hyaluronan
26	28	87.5	554	19	AAU850010	Mouse hyaluronate
27	28	87.5	554	21	AAU68493	Mouse hyaluronan s
28	28	87.5	554	24	ABP96030	Listeria monocytog
29	28	87.5	779	23	ABR48616	Drosophila melanog
30	28	87.5	814	22	ABR65317	Human secreted pro
31	27	84.4	46	21	AAU39407	Human ORFX protein
32	27	84.4	62	23	ABP01177	Human brain expres
33	27	84.4	165	22	AAU63373	Human polypeptide
34	27	84.4	229	23	ABP69846	Drosophila melanog
35	27	84.4	357	22	ABR70413	Human kinesi moto
36	27	84.4	357	23	AAO19019	Human kinesi moto
37	27	84.4	357	23	AAU74558	Human polypeptide
38	27	84.4	421	22	AAU41820	Human NOVI0b prote
39	27	84.4	511	23	ABG94648	Protein of NOVA 7
40	27	84.4	536	23	AAU04646	Human kinesi moto
41	27	84.4	563	23	AAO19018	Human kinesi moto
42	27	84.4	563	23	AAU74557	FLJ20354 f1s clone
43	27	84.4	811	23	ABP43909	Streptococcus poly
44	27	84.4	857	23	ABP26367	Streptococcus poly
45	27	84.4	876	23	ABP30087	Streptococcus poly

ALIGNMENTS

RESULT 1

AAE14548

ID AAE14548 standard; peptide; 7 AA.

XX AC AAE14548;

XX AC AAE14548; (first entry)

XX DT 17-MAY-2002 (first entry)

XX DE Human alpha-synuclein aggregation inhibitor #3.

XX KW Alpha-synuclein; inhibitor; neurodegenerative disease; Lewy body;

XX KW Parkinson's disease; Alzheimer's disease; diffuse Lewy body disease;

XX KW multiple system atrophy; Hallervorden-Spatz disease; human.

XX OS Homo sapiens.

XX PN WO200204482-A1

XX PD 17-JAN-2002.

XX PF 06-JUL-2001; 2001WO-US21379.

XX PR 07-JUL-2000; 2000US-217319P.

XX PR 28-MAR-2001; 2001US-279199P.

XX PA (PANA-) PANACEA PHARM INC.

XX PI Wolozin B, Ostretova-Golts N, Lebowitz MS;

XX DR WPI; 2002-179695/23.

XX PT Determination of an agent capable of inhibiting aggregation of alpha

PT synuclein useful for treating a neurodegenerative disease involves  
 PT determining aggregation of alpha synuclein in the presence of exogenous  
 PT iron or copper -  
 PS  
 PS Claim 40; Page 37; 52pp; English.  
 CC  
 CC The invention relates to screening of inhibitors of alpha-synuclein  
 CC aggregation in the presence of exogenous iron or copper. The inhibitors  
 CC are magnesium and alpha-synuclein binding peptides, which are  
 CC useful for treating neurodegenerative disease that involves  
 CC the formation of Lewy bodies e.g. Parkinson's disease (PD), Alzheimer's  
 CC disease (AD), diffuse Lewy body disease, mixed AD-PD, multiple system  
 CC atrophy and Hallervorden-Spatz disease. The present sequence is a  
 CC peptide that binds to C-terminal portion of human alpha-synuclein and  
 CC inhibits its aggregation.  
 CC  
 SQ Sequence 7 AA;  
 Query Match 100.0%; Score 32; DB 23; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ATINKSL 7  
 DB 1 ATINKSL 7  
 RESULT 2  
 AAU39185  
 ID AAU39185 standard; Protein; 74 AA.  
 AC AAU39185;  
 XX  
 DT 13-FEB-2002 (first entry)  
 DE  
 DE Propionibacterium acnes immunogenic protein #81.  
 KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 KW dermatological; osteopathic; neuroprotectant.  
 XX  
 OS Propionibacterium acnes.  
 XX  
 XX WO200181581-A2.  
 XX  
 XX 01-NOV-2001.  
 XX  
 XX 20-APR-2001; 2001WO-US12865.  
 XX  
 XX 21-APR-2000; 2000US-199047P.  
 XX 02-JUN-2000; 2000US-208841P.  
 XX 07-JUL-2000; 2000US-216747P.  
 XX  
 XX (CORI-) CORIXA CORP.  
 XX  
 XX Skeiky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
 XX L'Maisonneuve J, Zhang Y, Jen S, Carter D;  
 XX  
 XX WPI; 2001-616774/71.  
 XX N-PSDB; AAS959506.  
 XX  
 XX Propionibacterium acnes polypeptides and nucleic acids useful for  
 XX vaccinating against and diagnosing infections, especially useful for  
 XX treating acne vulgaris -  
 XX  
 XX Example 1; SEQ ID No 380; 1069pp; English.  
 PS  
 PS Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
 CC polypeptides. The proteins and their associated DNA sequences are used in  
 CC the treatment, prevention and diagnosis of medical conditions caused by  
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.

CC P. acnes is also involved in infections of bone, joints and the central  
 CC nervous system, however it is particularly involved in the inflammatory  
 CC lesions associated with acne vulgaris. A method for detecting the  
 CC presence or absence of P. acnes in a patient comprises contacting a  
 CC sample with a binding agent that binds to the proteins of the invention  
 CC and determining the amount of bound protein in the sample. The  
 CC polypeptides may be used as antigens in the production of antibodies  
 CC specific for P. acnes proteins. These antibodies can be used to  
 CC downregulate expression and activity of P. acnes polypeptides and  
 CC therefore treat P. acnes infections. The antibodies may also be used as  
 CC diagnostic agents for determining P. acnes presence, for example, by  
 CC enzyme linked immunosorbent assay (ELISA).  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 SQ Sequence 74 AA;  
 Query Match 93.8%; Score 30; DB 22; Length 74;  
 Best Local Similarity 85.7%; Pred. No. 24;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ATINKSL 7  
 DB 49 ATINKSL 55  
 RESULT 3  
 AAO02587  
 ID AAO02587 standard; Protein; 139 AA.  
 XX  
 AC AAO02587;  
 XX  
 DT 06-NOV-2001 (first entry)  
 DE  
 DE Human polypeptide SEQ ID NO 16479.  
 XX  
 XX Human; cytokines; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorders; arthritis; inflammation.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200164835-A2.  
 XX  
 XX 07-SEP-2001.  
 XX  
 XX 26-FEB-2001; 2001WO-US04927.  
 XX  
 XX 28-FEB-2000; 2000US-0515126.  
 XX 18-MAY-2000; 2000US-0577409.  
 XX  
 XX (HYSE-) HYSEQ INC.  
 XX  
 XX Tang YT, Liu C, Drmanac RT;  
 XX  
 XX WPI; 2001-514838/56.  
 XX N-PSDB; AAI82518.  
 XX  
 XX Isolated nucleic acids and polypeptides, useful for preventing  
 XX diagnosing and treating e.g. leukaemia, inflammation and immune  
 XX disorders -  
 XX  
 XX Claim 20; SEQ ID NO 16479; 1399pp + Sequence Listing; English.  
 PS  
 PS The invention relates to human polynucleotides (AAI79941-AAI93841) and  
 CC the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating



CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 139 AA;  
 Query Match 87.5%; Score 28; DB 22; Length 139;  
 Best Local Similarity 71.4%; Pred. No. 1.4e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ATINKSL 7  
 ||:||||:  
 Db 20 ATVAKAL 26

RESULT 4  
 ABG56567  
 ID ABG56567 standard; Peptide; 149 AA.  
 XX AC ABG56567;  
 XX DT 25-FEB-2003 (first entry)  
 XX DE Human liver peptide, SEQ ID No 35215.  
 XX KW Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;  
 XX KW hypercholesterolaemia; coronary heart disease.  
 XX OS Homo sapiens.  
 XX PN WO200157273-A2.  
 XX PD 09-AUG-2001.  
 XX PF 30-JAN-2001; 2001WO-US00664.  
 XX PR 04-FEB-2000; 2000US-0180312.  
 XX PR 26-MAY-2000; 2000US-0207456.  
 XX PR 30-JUN-2000; 2000US-0608408.  
 XX PR 03-AUG-2000; 2000US-0632366.  
 XX PR 21-SEP-2000; 2000US-0234687.  
 XX PR 27-SEP-2000; 2000US-0236359.  
 XX PR 04-OCT-2000; 2000GB-0024263.  
 XX PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX DR WPI; 2001-488898/53.  
 XX PT Human genome-derived single exon nucleic acid probes useful for  
 XX PT analysing gene expression in human adult liver -

Claim 27; SEQ ID No 35215; 659pp; English.  
 The invention relates to a single exon nucleic acid probe (SENP) (I) for  
 measuring human gene expression in a sample derived from human adult  
 liver, comprising one of 13109 defined nucleotide sequences given in the  
 specification (or complements/ fragments). The probe hybridises at high  
 stringency to a nucleic acid molecule expressed in the human adult  
 liver. (I) may be used for predicting, measuring and displaying gene  
 expression in samples derived from human adult liver. The genes  
 identified may be involved in genetic liver diseases such as cirrhosis,  
 hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which  
 is associated with coronary heart disease. ABG47348-ABG59930 represent  
 human liver single exon encoded peptides of the invention.  
 Note: The sequence information for this patent does not appear in the  
 printed specification but was obtained in electronic format directly  
 from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 149 AA;  
 Query Match 87.5%; Score 28; DB 22; Length 149;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 TINKSL 7  
 |||||  
 Db 108 TINKSL 113  
 RESULT 5  
 ABB41125  
 ID ABB41125 standard; Peptide; 149 AA.  
 XX AC ABB41125;  
 XX DT 04-FEB-2002 (first entry)  
 XX DE Peptide #8631 encoded by human foetal liver single exon probe.  
 XX KW Human; foetal liver; gene expression; single exon nucleic acid probe.  
 XX OS Homo sapiens.  
 XX PN WO200157277-A2.  
 XX PD 09-AUG-2001.  
 XX PF 30-JAN-2001; 2001WO-US00669.  
 XX PR 04-FEB-2000; 2000US-0180312.  
 XX PR 26-MAY-2000; 2000US-0207456.  
 XX PR 30-JUN-2000; 2000US-0608408.  
 XX PR 03-AUG-2000; 2000US-0632366.  
 XX PR 21-SEP-2000; 2000US-0234687.  
 XX PR 27-SEP-2000; 2000US-0236359.  
 XX PR 04-OCT-2000; 2000GB-0024263.  
 XX PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX DR WPI; 2001-483447/52.  
 XX PT Human genome-derived single exon nucleic acid probes useful for  
 XX PT analysing gene expression in human foetal liver -

Claim 27; SEQ ID No 33760; 639pp + sequence listing; English.  
 The invention relates to a single exon nucleic acid probe for  
 measuring human gene expression in a sample derived from human foetal  
 liver. The single exon nucleic acid probes may be used for predicting,  
 measuring and displaying gene expression in samples derived from human  
 foetal liver. The present sequence is a peptide encoded by a single exon  
 nucleic acid probe of the invention.  
 Note: The sequence data for this patent did not form part of the  
 printed specification, but was obtained in electronic format directly  
 from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 149 AA;  
 Query Match 87.5%; Score 28; DB 22; Length 149;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 TINKSL 7  
 |||||  
 Db 108 TINKSL 113  
 RESULT 6

ABB25168  
ID ABB25168 standard; Protein; 149 AA.  
AC ABB25168;  
XX  
DT 23-JAN-2002 (first entry)  
XX  
DE Protein #7167 encoded by probe for measuring heart cell gene expression.  
XX  
KW Human; gene expression; heart; microarray; vascular system;  
KW cardiovascular disease; hypertension; cardiac arrhythmia;  
KW congenital heart disease.  
XX  
OS Homo sapiens.  
XX  
PN WO200157274-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US00666.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
DR WPI; 2001-488999/53.  
XX  
XX Single exon nucleic acid probes for analyzing gene expression in human  
PT hearts -  
XX  
PS Claim 15; SEQ ID No 26938; 530pp; English.  
XX  
CC The present invention relates to single exon nucleic acid probes for  
CC measuring human gene expression in a sample derived from human heart (see  
CC ABA21535-ABA41305). The present sequence is a protein encoded by one such  
CC probe. The probes may be used for predicting, measuring and displaying  
CC gene expression in samples derived from the human heart via microarrays.  
CC By measuring gene expression, the probes are useful for predicting,  
CC diagnosing, grading, staging, monitoring and prognosing diseases of the  
CC human heart and vascular system e.g. cardiovascular disease,  
CC hypertension, cardiac arrhythmias and congenital heart disease.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 149 AA;  
XX  
Query Match 87.5%; Score 28; DB 22; Length 149;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 TINKSL 7  
Db 108 TINKSL 113  
RESULT 7  
AAM61981  
ID AAM61981 standard; Protein; 149 AA.  
XX  
AC AAM61981;  
XX  
DT 05-NOV-2001 (first entry)  
XX  
DE Human brain expressed single exon probe encoded protein SEQ ID NO: 34085.  
XX  
Human; brain expressed exon; gene expression analysis; probe;  
microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
epilepsy; cancer.  
XX  
OS Homo sapiens.  
XX  
PN WO200157275-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US00667.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
DR WPI; 2001-483446/52.  
XX  
XX Single exon nucleic acid probes for analyzing gene expression in human  
PT brains -  
XX  
PS Example 4; SEQ ID NO: 34086; 650pp + Sequence listing; English.  
XX  
CC The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC brain. They can be used to measure gene expression in brain cell samples,  
CC which may enable the diagnosis and improved treatment of nervous system  
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
CC epilepsy and cancers. The present sequence is a protein encoded by one of  
CC the probes of the invention.  
XX  
SQ Sequence 149 AA;  
XX  
Query Match 87.5%; Score 28; DB 22; Length 149;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 TINKSL 7  
Db 108 TINKSL 113  
RESULT 8  
AAM74785  
ID AAM74785 standard; Protein; 149 AA.  
XX  
AC AAM74785;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human bone marrow expressed probe encoded protein SEQ ID NO: 35091.  
XX  
KW Human; bone marrow expressed exon; gene expression analysis; probe;  
KW microarray; cancer; leukaemia; lymphoma; myeloma.  
XX  
OS Homo sapiens.  
XX  
PN WO200157276-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US00668.  
XX  
PR 04-FEB-2000; 2000US-0190312.

PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX (MOLE-) MOLECULAR DYNAMICS INC.  
 PA Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX WPI; 2001-488900/53.  
 XX Human genome-derived single exon nucleic acid probes useful for  
 PT analyzing gene expression in human bone marrow -  
 XX Example 4; SEQ ID NO: 35091; 658pp + Sequence Listing; English.  
 XX The present invention provides a number of single exon nucleic acid  
 CC probes which are derived from genomic sequences expressed in the human  
 CC bone marrow. They can be used to measure gene expression in bone marrow  
 CC samples, which may enable the improved diagnosis and treatment of cancers  
 CC such as lymphoma, leukaemia and myeloma. The present sequence is a  
 CC protein encoded by one of the probes of the invention.  
 XX SQ Sequence 149 AA;  
 Query Match 87.5%; Score 28; DB 22; Length 149;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02; Indels 0; Gaps 0;  
 Matches 6; Conservative 0; Mismatches 0;  
 QY 2 TINKSL 7  
 Db 108 TINKSL 113  
 RESULT 9  
 AAM34901  
 ID AAM34901 standard; Protein; 149 AA.  
 XX AC AAM34901;  
 XX 17-OCT-2001 (first entry)  
 DT Peptide #8938 encoded by probe for measuring placental gene expression.  
 DE Probe; microarray; human; placenta; antenatal diagnosis;  
 KW genetic disorder.  
 KW Homo sapiens.  
 OS WO200157272-A2.  
 FN 09-AUG-2001.  
 PD 30-JAN-2001; 2001WO-US00663.  
 PF 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX (MOLE-) MOLECULAR DYNAMICS INC.  
 PA Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX WPI; 2001-488997/53.  
 XX Human genome-derived single exon nucleic acid probes useful for  
 PT analyzing gene expression in human placenta -

XX Claim 27; SEQ ID No 35170; 654pp; English.  
 XX The present invention relates to single exon nucleic acid probes (SENP:  
 CC see A131315-AA157546). The present sequence is a peptide encoded by one  
 CC such probe. The probes are useful for producing a microarray for  
 CC predicting, measuring and displaying gene expression in samples derived  
 CC from human placenta. The probes are useful for antenatal diagnosis of  
 CC human genetic disorders.  
 XX SQ Sequence 149 AA;  
 Query Match 87.5%; Score 28; DB 22; Length 149;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02; Indels 0; Gaps 0;  
 Matches 6; Conservative 0; Mismatches 0;  
 QY 2 TINKSL 7  
 Db 108 TINKSL 113  
 RESULT 10  
 ABG44576  
 ID ABG44576 standard; Peptide; 149 AA.  
 XX AC ABG44576;  
 XX 19-AUG-2002 (first entry)  
 DT Human peptide encoded by genome-derived single exon probe SEQ ID 34241.  
 DE Human; single exon probe; asthma; lung cancer; COPD; ILD;  
 KW chronic obstructive pulmonary disease; interstitial lung disease;  
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;  
 KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;  
 KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;  
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;  
 KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
 KW primary ciliary dyskinesia; pulmonary hypertension;  
 KW hyaline membrane disease.  
 XX Homo sapiens.  
 OS WO200186003-A2.  
 FN 15-NOV-2001.  
 PD 30-JAN-2001; 2001WO-US00665.  
 PF 04-FEB-2000; 2000US-180312P.  
 PR 26-MAY-2000; 2000US-207456P.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-234687P.  
 PR 27-SEP-2000; 2000US-236359P.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX (MOLE-) MOLECULAR DYNAMICS INC.  
 PA Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX WPI; 2002-114183/15.  
 XX Spatially-addressable set of single exon nucleic acid probes, used to  
 PT measure gene expression in human lung samples -  
 XX Claim 27; SEQ ID No 34241; 634pp; English.  
 XX The invention relates to a spatially-addressable set of single exon  
 CC nucleic acid probes for measuring gene expression in a sample derived  
 CC from human lung comprising single exon nucleic acid probes having one of  
 CC 12614 nucleic acid sequences mentioned in the specification, or their  
 CC complements or the 12387 open reading frames derived from the 12614

CC probes. Also included are a microarray comprising the novel set of  
 CC probes; the novel set of probes which hybridise at high stringency to a  
 CC nucleic acid expressed in the human lung; measuring gene expression in a  
 CC sample derived from human lung, comprising (a) contacting the array with  
 CC a collection of detectably labeled nucleic acids derived from human lung  
 CC mRNA, and (b) measuring the label detectably bound to each probe of  
 CC the array; identifying exons in a eukaryotic genome, comprising  
 CC (a) algorithmically predicting at least one exon from genomic sequences  
 CC of the eukaryote; and (b) detecting specific hybridisation of detectably  
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
 CC having a fragment identical to the predicted exon, the probe is included  
 CC in the above mentioned microarray; assigning exons to a single gene,  
 CC comprising (a) identifying exons from genomic sequence by the method  
 CC above and (b) measuring the expression of each of the exons in several  
 CC tissues and/or cell types using hybridisation to a single exon  
 CC microarrays having a probe with the exon, where a common pattern of  
 CC expression of the exons in the tissues and/or cell types indicates that  
 CC the exons should be assigned to a single gene; a peptide comprising one  
 CC of 12011 sequences, mentioned in the specification, or encoded by the  
 CC probes/open reading frames (ORF). The probes are used for gene  
 CC expression analysis, and for identifying exons in a gene, particularly  
 CC using human lung derived mRNA and for the study of lung diseases  
 CC such as asthma, lung cancer, chronic obstructive pulmonary disease  
 CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary  
 CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,  
 CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary  
 CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,  
 CC pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic  
 CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension  
 CC and hyaline membrane disease. The present sequence is a peptide/protein  
 CC encoded by a single exon probe of the invention.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pot\_sequences.

XX Sequence 149 AA;  
 Query Match 87.5%; Score 28; DB 23; Length 149;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TINKSL 7  
 DB 108 TINKSL 113  
 |||||

RESULT 11  
 AA50013  
 ID AAW50013 standard; peptide; 190 AA.  
 XX AAW50013;  
 AC AAW50013;  
 DT 11-JUN-1998 (first entry)  
 DE Human hyaluronan synthase-3 partial peptide.  
 XX Human; hyaluronan synthase-3; HS-3; hyaluronan;  
 KW hyaluronic acid.  
 XX Homo sapiens.  
 OS  
 XX WO9800551-A2.  
 PN 08-JAN-1998.  
 PD  
 XX 03-JUL-1997; 97WO-US11761.  
 PF  
 XX 05-MAR-1997; 97US-0812008.  
 PR 03-JUL-1996; 96US-0675499.  
 PR  
 XX (MAYO-) MAYO FOUNDATION.  
 PA  
 XX

PI Augustine ML, McDonald JA, Spicer AP;  
 XX WPI; 1998-086976/08.  
 DR  
 XX Nucleic acid encoding hyaluronan synthase-2 or -3 or their active  
 PT fragments - used for treating arthritis, in wound repair, for  
 PT delivering drugs etc.  
 XX  
 PS Claim 14; Page 74; 115pp; English.  
 XX  
 CC The present sequence is a human hyaluronan synthase-3 (HS-3)  
 CC partial peptide, which can be used to alter hyaluronan, i.e.  
 CC hyaluronic acid (HA), production, or for in vitro synthesis of HA.  
 CC HA can be used to treat arthritis or perforated ear drums, protect  
 CC eyes during surgery, deliver drugs, prevent post-operative scarring  
 CC or adhesion formation, heal wounds and prevent restenosis following  
 CC balloon angioplasty. Modulation of HA production in vivo may be  
 CC useful in, e.g. Graves disease, mesothelioma, Wilm's tumour and  
 CC oedema associated with inflammation of lung and kidney, all of  
 CC which are associated with elevated levels of HA in tissues or  
 CC serum.  
 XX Sequence 190 AA;  
 Query Match 87.5%; Score 28; DB 19; Length 190;  
 Best Local Similarity 100.0%; Pred. No. 2e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATKINS 6  
 DB 185 ATKINS 190  
 |||||

RESULT 12  
 AAU27524  
 ID AAU27524 standard; Protein; 199 AA.  
 XX AAU27524;  
 AC  
 XX 18-DEC-2001 (first entry)  
 DT  
 XX Human G-Protein Coupled Receptor (GPCR) polypeptide #52.  
 DE  
 XX Human; G-protein coupled receptor; GPCR; mental disorder; schizophrenia;  
 KW attention deficit disorder; anxiety; depression; bipolar disorder;  
 KW neurological disorder; Huntington's disease; dementia; obesity; anorexia;  
 KW metabolic disorder; Parkinson's disease; Tourette's syndrome; thrombosis;  
 KW type 2 diabetes; cardiovascular disorder; myocardial infarction; cancer;  
 KW cardiomyopathy; atherosclerosis; human immunodeficiency virus; HIV;  
 KW viral infection; immunostimulant; neuroleptic; nootropic; tranquiliser;  
 KW antidepressant; anorectic; gene therapy; antiviral; cardiant; alopecia;  
 KW neuroprotective; cyostatic; antiparkinsonian; antidiabetic; psoriasis;  
 KW hyperproliferative disorder; hormonal disorder; inflammatory disorder;  
 KW Crohn's disease.  
 XX  
 XX Homo sapiens.  
 OS  
 XX WO200162798-A2.  
 PN 30-AUG-2001.  
 PD  
 XX 23-FEB-2001; 2001WO-US05985.  
 PF  
 XX 25-FEB-2000; 2000US-0184939.  
 PR 25-FEB-2000; 2000US-0184940.  
 PR 25-FEB-2000; 2000US-0184941.  
 PR 25-FEB-2000; 2000US-0184942.  
 PR 25-FEB-2000; 2000US-0185042.  
 PR 25-FEB-2000; 2000US-0185044.  
 PR 25-FEB-2000; 2000US-0185208.  
 PR 13-MAR-2000; 2000US-0184843.  
 PR  
 XX (PHAA ) PHARMACIA & UPJOHN CO.  
 PA



PR 18-AUG-1999; 99US-0149499.  
 PR 28-SEP-1999; 99US-0156503.  
 (GEHU-) GERMAN HUMAN GENOME PROJECT.  
 XX Wiemann S;  
 XX WPI; 2001-327840/34.  
 XX Nucleic acids having the sequences of clones isolated from libraries of  
 PT different human tissues, useful in recombinant DNA methodologies -  
 XX Example III; Page 743; 1095pp; English.  
 XX This invention describes novel polynucleotides and polypeptides isolated  
 CC from human cDNA libraries which can be used for gene therapy or in  
 CC vaccines. The polynucleotides of the invention and antibodies encoded by  
 CC them may be used in the prevention, diagnosis and treatment of diseases  
 CC associated with inappropriate polypeptide expression. The products of the  
 CC invention may also be used to identify modulators of expression and  
 CC activity and to down regulate expression and activity. The antibodies of  
 CC the invention may also be used as diagnostic agents for detecting the  
 CC presence of polypeptides in samples. This sequence represents a homologue  
 CC of a polypeptide described in the disclosure of the invention.  
 XX  
 SQ Sequence 339 AA;  
 Query Match 87.5%; Score 28; DB 22; Length 339;  
 Best Local Similarity 85.7%; Pred. NO. 3.8e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ATINKSL 7  
 :|||||  
 Db 266 ASINKSL 272  
 RESULT 15  
 ABG24679  
 ID ABG24679 standard; Protein; 395 AA.  
 XX AC ABG24679;  
 XX 18-FEB-2002 (first entry)  
 XX Novel human diagnostic protein #24670.  
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX Homo sapiens.  
 XX WO200175067-A2.  
 XX 11-OCT-2001.  
 XX 30-MAR-2001; 2001WO-US08631.  
 XX 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX (HYSE-) HYSEQ INC.  
 XX Drmanac RT, Liu C, Tang YT;  
 XX WPI; 2001-639362/73.  
 DR N-PSDB; AAS88866.  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX Claim 20; SEQ ID No 55038; 103pp; English.  
 PS

XX The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (III). (III) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 395 AA;  
 Query Match 87.5%; Score 28; DB 22; Length 395;  
 Best Local Similarity 100.0%; Pred. NO. 4.5e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 TINKSL 7  
 :|||||  
 Db 341 TINKSL 346

Search completed: February 11, 2004, 17:02:46  
 Job time : 33.25 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 16:52:34 ; Search time 24.5833 Seconds  
(without alignments)  
73.479 Million cell updates/sec

Title: US-09-901-187C-4  
Perfect score: 34  
Sequence: 1 RRRGNMAI 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL 23:\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rv.\*
- 16: sp\_bacteriaph.\*
- 17: sp\_archaeal.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	97.1	232	12 Q9DSP1	Q9dsp1 saint croix
2	33	97.1	320	16 Q9CBL8	Q9cbl8 mycobacteri
3	32	94.1	267	16 Q9R130	Q9r130 streptomyc
4	30	88.2	263	16 Q9CBL2	Q9cbl2 mycobacteri
5	30	88.2	321	16 Q915P3	Q915p3 pseudomonas
6	30	88.2	345	10 Q94JL0	Q94jl0 oryza sativ
7	30	88.2	349	17 Q8TVK4	Q8tvk4 methanopyru
8	30	88.2	456	10 Q9AYF7	Q9ayf7 oryza sativ
9	30	88.2	465	5 Q81G01	Q81g01 caenorhabdi
10	30	88.2	567	5 Q21463	Q21463 caenorhabdi
11	30	88.2	684	16 Q8G785	Q8g785 bifidobacte
12	30	88.2	695	5 Q9NJB3	Q9njb3 drosophila
13	30	88.2	695	5 Q95WB5	Q95wb5 drosophila
14	30	88.2	695	5 Q9NJB1	Q9njb1 drosophila
15	30	88.2	735	5 Q8SXC4	Q8sxc4 drosophila
16	30	88.2	943	16 Q9Z5A9	Q9z5a9 streptomyc

17	30	88.2	1280	2 Q9F6X9	Q9f6x9 chloroflexu
18	30	88.2	1335	5 Q9VFZ9	Q9vfz9 drosophila
19	30	88.2	3429	5 Q8IBP1	Q8ibp1 plasmodium
20	29	85.3	62	16 Q8UA65	Q8uae5 agrobacteri
21	29	85.3	137	5 Q27749	Q27749 psammecchinu
22	29	85.3	171	16 Q92295	Q92295 rhizobium m
23	29	85.3	172	10 Q9FSL4	Q9fsl4 oryza sativ
24	29	85.3	218	2 Q9FIQ2	Q9fiq2 thermus the
25	29	85.3	246	16 Q8YKT0	Q8ykt0 anabaena sp
26	29	85.3	272	16 Q53572	Q53572 mycobacteri
27	29	85.3	298	16 Q9RVH1	Q9rvh1 deinococcus
28	29	85.3	359	16 Q9HW62	Q9hw62 pseudomonas
29	29	85.3	361	16 Q9RVH2	Q9rvh2 deinococcus
30	29	85.3	437	16 Q92RP9	Q92rp9 rhizobium m
31	29	85.3	442	16 Q8FLZ4	Q8flz4 corynebacte
32	29	85.3	469	16 Q98K12	Q98k12 rhizobium l
33	29	85.3	501	12 Q98111	Q98111 gallid herp
34	29	85.3	695	5 Q9U8E0	Q9u8e0 hirtodrosop
35	29	85.3	773	16 Q92JW6	Q92jw6 rhizobium m
36	29	85.3	784	16 Q8P5M0	Q8p5m0 xanthomonas
37	29	85.3	912	10 Q91QJ5	Q91qj5 arabidopsis
38	29	85.3	919	10 Q9SNW1	Q9snw1 arabidopsis
39	29	85.3	919	10 Q9LNR3	Q9lnr3 arabidopsis
40	29	85.3	922	10 Q9FSE5	Q9fse5 oryza sativ
41	28	82.4	60	16 Q8YJY9	Q8yjy9 anabaena sp
42	28	82.4	94	10 Q943U3	Q943u3 oryza sativ
43	28	82.4	114	10 Q8LMP2	Q8lmp2 oryza sativ
44	28	82.4	123	10 Q8S6P4	Q8s6p4 oryza sativ
45	28	82.4	139	10 Q8GSX2	Q8gsx2 oryza sativ

ALIGNMENTS

RESULT 1

Q9DSP1 PRELIMINARY; PRT; 232 AA.  
ID Q9DSP1  
AC Q9DSP1;  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE VP6.  
OS Saint Croix river virus.  
OC Viruses; dsRNA viruses; Reoviridae; Orbivirus.  
OX NCBI\_TaxID=104581;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=2115717; PubMed=11257184;  
RA Attoui H., Stirling J.M., Munderloh U.G., Billoir F., Brookes S.M.,  
RA Burroughs J.N., de Micco P., Mertens P.C., de Lamballerie X.;  
RT "Complete sequence characterization of the genome of the St Croix  
RT River virus, a new orbivirus isolated from cells of ixodes  
RT scapularis.";  
RL J. Gen. Virol. 82:795-804 (2001).  
DR EMBL; AF145406; AAG34265.1; -.  
SQ SEQUENCE 232 AA; 25775 MW; 4B99FC439CB583F4 CRC64;

Query Match 97.1%; Score 33; DB 12; Length 232;  
Best Local Similarity 85.7%; Pred. No. 23;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRRGNMAI 7

Db 26 RRRGNMAV 32

RESULT 2

Q9CBL8 PRELIMINARY; PRT; 320 AA.  
ID Q9CBL8  
AC Q9CBL8;  
DT 01-JUN-2001 (TREMBlrel. 17, Created)  
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
DT 01-OCT-2001 (TREMBlrel. 18, Last annotation update)

DE Hypothetical protein ML1809.  
 GN ML1809.  
 OS Bacterium leprae.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1769;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TN;  
 RX MEDLINE=21128732; PubMed=11234002;  
 RA Cole S.T., Eiglmeyer K., Parkhill J., James K.D., Thomson N.R.,  
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,  
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,  
 RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,  
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,  
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,  
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,  
 RA Barrell B.G.;  
 RT "Massive gene decay in the leprosy bacillus."  
 RL Nature 409:1007-1011(2001).  
 DR EMBL; AL583923; CAC30762.1; -.  
 DR Leproma; ML1809; -.  
 DR InterPro; IPR002881; DUF58.  
 DR Pfam; PF01882; DUF58; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 320 AA; 34944 MW; FA405629388219CD CRC64;

Query Match 97.1%; Score 33; DB 16; Length 320;  
 Best Local Similarity 85.7%; Pred. No. 32;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRRGMAI 7  
 Db 197 RRRGMV 203  
 |||||:  
 197 RRRGMV 203

RESULT 3  
 Q9RI30 PRELIMINARY; PRT; 267 AA.  
 ID Q9RI30;  
 AC Q9RI30;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Putative nitrate reductase delta chain.  
 GN NARJ2 OR SC00218 OR SCJ12.30.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Streptomycineae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Murphy L., Harris D.;  
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;  
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RX MEDLINE=97000351; PubMed=8843436;  
 RA Redenbach M., Kieser H.M., Denapate D., Eichner A., Cullum J.,  
 RA Kinashi H., Hopwood D.A.;  
 RT "A set of ordered cosmids and a detailed genetic and physical map for  
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
 RL Mol. Microbiol. 21:77-96(1996).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2) / ML145;  
 RX MEDLINE=21996410; PubMed=12000953;

RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,  
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,  
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
 RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,  
 RA Rabinowitch E., Rajandream M.A., Rutherford K., Rutter S.,  
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
 RA Warren T., Wietzorek A., Woodward J., Barrell B.G., Parkhill J.,  
 RA Hopwood D.A.;  
 RT "Complete genome sequence of the model actinomycete Streptomyces  
 RT coelicolor A3(2)."  
 RL Nature 417:141-147(2002).  
 DR EMBL; AL939104; CABS3442.1; -.  
 DR InterPro; IPR003765; Nitrate red del.  
 DR Pfam; PF02613; Nitrate red del; 1.  
 DR TIGRFAMs; TIGR00684; narJ; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 267 AA; 28720 MW; EAP80F543193D10 CRC64;

Query Match 94.1%; Score 32; DB 16; Length 267;  
 Best Local Similarity 85.7%; Pred. No. 44;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRRGMAI 7  
 Db 127 RRRGMAL 133  
 |||||:  
 127 RRRGMAL 133

RESULT 4  
 Q9CBJ2 PRELIMINARY; PRT; 263 AA.  
 ID Q9CBJ2;  
 AC Q9CBJ2;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)  
 DE Hypothetical membrane protein ML1918.  
 GN ML1918.  
 OS Mycobacterium leprae.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1769;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TN;  
 RX MEDLINE=21128732; PubMed=11234002;  
 RA Cole S.T., Eiglmeyer K., Parkhill J., James K.D., Thomson N.R.,  
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,  
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,  
 RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,  
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,  
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,  
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,  
 RA Barrell B.G.;  
 RT "Massive gene decay in the leprosy bacillus."  
 RL Nature 409:1007-1011(2001).  
 DR EMBL; AL583923; CAC30873.1; -.  
 DR Leproma; ML1918; -.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 263 AA; 27165 MW; 3E97C9FB3E65652A CRC64;

Query Match 88.2%; Score 30; DB 16; Length 263;  
 Best Local Similarity 71.4%; Pred. No. 1.2e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRRGMAI 7  
 Db 17 RRRGLAV 23  
 |||||:  
 17 RRRGLAV 23

RESULT 5  
 Q915P3 PRELIMINARY; PRT; 321 AA.  
 ID Q915P3



Q915P3;  
 AC 01-MAR-2001 (T-EMBLrel. 16, Created)  
 DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)  
 DT 01-OCT-2002 (T-EMBLrel. 22, Last annotation update)  
 DE Probable type II secretion system protein.  
 GN PA0682.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 OC Pseudomonadaceae; Pseudomonas.  
 OX NCBI\_TaxID=287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 15692 / PA01;  
 RX MEDLINE=20437337; PubMed=10994043;  
 RA Stover C.K., Pham C.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
 RA Garber R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,  
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
 RA Smith K.L., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.,  
 RA "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
 RT opportunistic pathogen."  
 RL Nature 406:959-964(2000).  
 DR EMBL; AE004503; AAG04071.1; --  
 DR InterPro; IPR005628; GspK.  
 DR Pfam; PF03934; GspK; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 321 AA; 36042 MW; 0FD4172DA21E9136 CRC64;

Query Match 88.2%; Score 30; DB 16; Length 321;  
 Best Local Similarity 85.7%; Pred. No. 1.5e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RRRGMWAI 7  
 Db 5 RRRGMWAI 11

RESULT 6  
 Q94JL0  
 ID Q94JL0 PRELIMINARY; PRT; 345 AA.  
 AC Q94JL0;  
 DT 01-DEC-2001 (T-EMBLrel. 19, Created)  
 DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)  
 DT 01-OCT-2002 (T-EMBLrel. 22, Last annotation update)  
 DE P0434B04.13 protein.  
 GN P0434B04.13.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Nipponbare;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC  
 RT clone:P0434B04."  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF002540; BAB43992.1; --  
 DR Granele; Q94JL0; --  
 SQ SEQUENCE 345 AA; 38547 MW; BEAE588ACF9FC9A0 CRC64;

Query Match 88.2%; Score 30; DB 10; Length 345;  
 Best Local Similarity 71.4%; Pred. No. 1.6e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RRRGMWAI 7  
 Db 66 RRRGLAV 72

RESULT 7

Q8TVK4  
 ID Q8TVK4 PRELIMINARY; PRT; 349 AA.  
 AC Q8TVK4;  
 DT 01-JUN-2002 (T-EMBLrel. 21, Created)  
 DT 01-JUN-2002 (T-EMBLrel. 21, Last sequence update)  
 DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)  
 DE Nitrogenase molybdenum-iron subunit.  
 GN NIFD OR KX1385.  
 OS Methanopyrus kandleri.  
 OC Archaea; Euryarchaeota; Methanopyri; Methanopyraceae;  
 OC Methanopyrus.  
 OX NCBI\_TaxID=2320;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AV19 / DSM 6324 / JCM 9639;  
 RX MEDLINE=21927647; PubMed=11930014;  
 RA Slesarev A.I., Mezhevaya K.V., Makarova K.S., Polushin N.N.,  
 RA Shcherbinina O.V., Shakhova V.V., Belova G.I., Aravind L.,  
 RA Natale D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,  
 RA Malykh A.G., Koonin E.V., Kozlovkin S.A.;  
 RA "The complete genome of hyperthermophile Methanopyrus kandleri AV19  
 RT and monophyly of archaeal methanogens."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).  
 DR EMBL; AE010431; AAM02598.1; --  
 DR InterPro; IPR000510; Oxred\_nitrogenase.  
 DR Pfam; PF00148; oxidored\_nitro; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 349 AA; 38156 MW; 628644FA14EC1CEE CRC64;

Query Match 88.2%; Score 30; DB 17; Length 349;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RRRGMWAI 6  
 Db 160 RRRGMWAI 165

RESULT 8  
 Q9AYF7  
 ID Q9AYF7 PRELIMINARY; PRT; 456 AA.  
 AC Q9AYF7;  
 DT 01-JUN-2001 (T-EMBLrel. 17, Created)  
 DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)  
 DT 01-OCT-2002 (T-EMBLrel. 22, Last annotation update)  
 DE Hypothetical 47.9 kDa protein.  
 GN OSJNBA0094J09.13.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC See L.H., Spiegel L.A., Nascimento L.U., de la Bastide M.,  
 RA Preston R.R., Huang E.N., Rodriguez M.A., Vil M.D., Baker J.P.,  
 RA Bahret A., Shah R.S., Miller B., Kirchoff K.A., King L., Toth K.,  
 RA O'Shaughnessy A., Dedhia N.N., McCombie W.R.;  
 RA "Genomic Sequence For Oryza sativa, Nipponbare strain, Chromosome X,  
 RT Clone OSJNBA0094J09, complete sequence."  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC McCombie W.R.;  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC McCombie W.R.;  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC Bal H.P., See L.H., Spiegel L.A., Nascimento L.U., de la Bastide M.,  
 RA Preston R.R., Huang E.N., Rodriguez M.A., Vil M.D., Baker J.P.,



Query Match 88.2%; Score 30; DB 16; Length 684;  
 Best Local Similarity 85.4%; Pred. No. 3.1e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRGMAI 7  
 Db 27 RRGMAI 33

## RESULT 12

Q9NJB3 Q9NJB3 PRELIMINARY; PRT; 695 AA.  
 ID Q9NJB3  
 AC Q9NJB3  
 DT 01-OCT-2000 (TREMELrel. 15, Created)  
 DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMELrel. 21, Last annotation update)  
 DE Xanthine dehydrogenase (Fragment).  
 GN XDH.  
 OS Drosophila erecta (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7220;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20119750; PubMed=10654254;  
 RA Rodriguez-Trelles F., Tarrío R., Ayala F.J.;  
 RT "Fluctuating mutation bias and the evolution of base composition in  
 RT Drosophila.";  
 RL J. Mol. Evol. 50:1-10(2000).  
 DR EMBL; AF169400; AAF31666.1; -  
 DR HSSP; P80457; 1FIQ.  
 DR FlyBase; FBgn024451; Dere\Xdh.  
 DR InterPro; IPR000674; Aldxan dh.C.  
 DR InterPro; IPR005107; CO deh flav.C.  
 DR InterPro; IPR005572; Euk Mb oxred.  
 DR Pfam; PF01315; Ald\_Xan dh\_C; 1.  
 DR Pfam; PF02738; Ald\_Xan dh\_C2; 1.  
 DR Pfam; PF03450; CO deh flav.C; 1.  
 DR PROSITE; PS00559; MOLYBDOPTERIN\_EUK; 1.  
 FT NON\_TER 1  
 FT NON\_TER 695  
 SQ SEQUENCE 695 AA; 76608 MW; 1BCOFFED5DEDA0D0 CRC64;

Query Match 88.2%; Score 30; DB 5; Length 695;  
 Best Local Similarity 71.4%; Pred. No. 3.2e-02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRGMAI 7  
 Db 489 RRGMAV 495

## RESULT 13

Q95WB5 Q95WB5 PRELIMINARY; PRT; 695 AA.  
 ID Q95WB5  
 AC Q95WB5  
 DT 01-DEC-2001 (TREMELrel. 19, Created)  
 DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)  
 DT 01-OCT-2002 (TREMELrel. 22, Last annotation update)  
 DE Xanthine dehydrogenase (Fragment).  
 GN XDH.  
 OS Drosophila ananassae (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7217;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21531852; PubMed=11675609;  
 RA Rodriguez-Trelles F., Tarrío R., Ayala F.J.;  
 RT "Xanthine dehydrogenase (XDH): episodic evolution of a 'neutral'  
 RT protein.";

J. Mol. Evol. 53:485-495(2001).  
 EMBL; AF345901; AAK97364.1; -  
 DR FlyBase; FBgn0024471; Aldxan dh.C.  
 DR InterPro; IPR000674; Aldxan dh.C.  
 DR InterPro; IPR005107; CO deh flav.C.  
 DR InterPro; IPR005572; Euk Mb oxred.  
 DR Pfam; PF01315; Ald\_Xan dh\_C; 1.  
 DR Pfam; PF02738; Ald\_Xan dh\_C2; 1.  
 DR Pfam; PF03450; CO deh flav.C; 1.  
 DR PROSITE; PS00559; MOLYBDOPTERIN\_EUK; 1.  
 FT NON\_TER 1  
 FT NON\_TER 695  
 SQ SEQUENCE 695 AA; 76582 MW; D4231111CA6E423A CRC64;

Query Match 88.2%; Score 30; DB 5; Length 695;  
 Best Local Similarity 71.4%; Pred. No. 3.2e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRGMAI 7  
 Db 489 RRGMAV 495

## RESULT 14

Q9NJB1 Q9NJB1 PRELIMINARY; PRT; 695 AA.  
 ID Q9NJB1  
 AC Q9NJB1  
 DT 01-OCT-2000 (TREMELrel. 15, Created)  
 DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMELrel. 21, Last annotation update)  
 DE Xanthine dehydrogenase (Fragment).  
 GN XDH.  
 OS Drosophila bifasciata (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7218;  
 RN [1]  
 RP SEQUENCE OF 1-324 FROM N.A.  
 RX MEDLINE=20119750; PubMed=10654254;  
 RA Rodriguez-Trelles F., Tarrío R., Ayala F.J.;  
 RT "Fluctuating mutation bias and the evolution of base composition in  
 RT Drosophila.";  
 RL J. Mol. Evol. 50:1-10(2000).  
 DR EMBL; AF169403; AAF31665.1; -  
 DR HSSP; P80457; 1FIQ.  
 DR FlyBase; FBgn0040472; Dbif\Xdh.  
 DR InterPro; IPR000674; Aldxan dh.C.  
 DR InterPro; IPR005107; CO deh flav.C.  
 DR InterPro; IPR005572; Euk Mb oxred.  
 DR Pfam; PF01315; Ald\_Xan dh\_C; 1.  
 DR Pfam; PF02738; Ald\_Xan dh\_C2; 1.  
 DR Pfam; PF03450; CO deh flav.C; 1.  
 DR PROSITE; PS00559; MOLYBDOPTERIN\_EUK; 1.  
 FT NON\_TER 1  
 FT NON\_TER 695  
 SQ SEQUENCE 695 AA; 76387 MW; 82EED64EAE3F561E CRC64;

Query Match 88.2%; Score 30; DB 5; Length 695;  
 Best Local Similarity 71.4%; Pred. No. 3.2e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRGMAI 7  
 Db 489 RRGMAV 495

## RESULT 15

Q8SXC4  
 ID Q8SXC4 PRELIMINARY; PRT; 735 AA.  
 AC Q8SXC4;  
 DT 01-JUN-2002 (TReMBLrel. 21, Created)  
 DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)  
 DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)  
 DE GH08847p.  
 GN RY OR CG7642.  
 CS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 CX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkely;  
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
 RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,  
 RA George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,  
 RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,  
 RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,  
 RA Ceiniker S.;  
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY094689; AAM11042.1; -.  
 DR FlyBase; FBgn0003308; ry.  
 DR InterPro; IPR000674; Aldxan dh C.  
 DR InterPro; IPR000572; Euk Mb-oxied.  
 DR Pfam; PF01315; Ald\_Xan dh C\_1  
 DR Pfam; PF02738; Ald\_Xan dh C2\_1  
 DR PROSITE; PS00559; MOUHDPTERIN\_EUK; 1.  
 SQ SEQUENCE 735 AA; 81002 MW; F6DE32CDDDE2B4BF CRC64;  
 Query Match 88.2%; Score 30; DB 5; Length 735;  
 Best Local Similarity 71.4%; Pred. No. 3.3e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RREGMAI 7  
 Db 399 RREGMAV 405

Search completed: February 11, 2004, 17:09:23  
 Job time : 26.5833 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 16:51:43 ; Search time 5.16667 Seconds  
(without alignments)  
63.714 Million cell updates/sec

Title: US-09-901-187C-4

Perfect score: 34

Sequence: 1 RRRGNMAY 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	33	97.1	169	1 YB80 MYCAV	O07394 mycobacteri
2	33	97.1	317	1 YB80 MYCTU	P71761 mycobacteri
3	30	88.2	70	1 L2MU ADE40	Q64858 human adeno
4	30	88.2	79	1 L2MU ADE02	P14269 human adeno
5	30	88.2	657	1 CTYP MYCTU	O53160 mycobacteri
6	30	88.2	716	1 SYEC SCHPO	O13775 schizosacch
7	30	88.2	1335	1 XDH DROME	P10351 drosophila
8	30	88.2	1342	1 XDH DROPS	P22811 drosophila
9	29	85.3	382	1 YH6 YEAST	P32793 saccharomyc
10	29	85.3	724	1 SYEC YEAST	P45655 saccharomyc
11	28	82.4	293	1 IC66 HUMAN	P52112 homo sapien
12	28	82.4	298	1 Y812 ARCFU	O29446 archaeoglob
13	28	82.4	305	1 YH6 ARCFU	P54592 bacillus su
14	28	82.4	363	1 ADA HUMAN	P09813 homo sapien
15	28	82.4	429	1 COB8 RHIME	Q92P48 rhizobium m
16	28	82.4	617	1 VP3B HUMAN	Q9H267 homo sapien
17	28	82.4	617	1 VP3B MOUSE	P59016 mus musculu
18	28	82.4	617	1 VP3B RAT	Q63616 rattus norv
19	28	82.4	640	1 GATE METKA	O87211 methanopyru
20	28	82.4	703	1 HS83 TRYB3	P12861 trypanosoma
21	28	82.4	1061	1 PRDA HUMAN	Q9H267 homo sapien
22	28	82.4	1065	1 SEC8 YEAST	P32855 saccharomyc
23	27	79.4	93	1 N36B SOYBN	Q02919 glycine max
24	27	79.4	194	1 3MGH MYCFO	Q9E6U9 mycobacteri
25	27	79.4	381	1 MRP MYCTU	O33225 mycobacteri
26	27	79.4	433	1 COB8 PSEDE	P21632 pseudomonas
27	27	79.4	453	1 SYS ARCFU	O28244 archaeoglob
28	27	79.4	453	1 VATE SULSO	Q9NW88 sulfolobus
29	27	79.4	495	1 ACCD MYCTU	P10561 mycobacteri
30	27	79.4	540	1 HUP2 CHLKE	Q39524 chlorella k
31	27	79.4	574	1 SEN3 HUMAN	Q9H414 homo sapien
32	27	79.4	777	1 NTRY AZOBR	P45675 azospirillum
33	27	79.4	859	1 LOX1 ARATH	Q06327 arabidopsis

34 27 79.4 1226 1 METH\_ECOLI  
35 27 79.4 1333 1 ADO\_MOUSE  
36 27 79.4 1333 1 ADO\_RAT  
37 27 79.4 1704 1 VITI\_FUNHE  
38 26 76.5 155 1 RL19\_AERPE  
39 26 76.5 173 1 Y057 MYCTU  
40 26 76.5 192 1 YBCC\_RHOCA  
41 26 76.5 197 1 DEF\_MYCLE  
42 26 76.5 197 1 DEF\_MYCTU  
43 26 76.5 276 1 Y938 TREPA  
44 26 76.5 299 1 ICE1\_SPOFR  
45 26 76.5 314 1 GSHE\_XYLFA

#### ALIGNMENTS

##### RESULT 1

YB80 MYCAV STANDARD; PRT; 169 AA.  
ID YB80 MYCAV STANDARD; PRT; 169 AA.  
AC O07394;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Hypothetical protein MAV169.  
OS Mycobacterium avium.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1764;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=GIR10;  
RX MEDLINE=98195739; PubMed=9534249;  
RA Labo M., Gusbetti L., de Rossi E., Speziale P., Riccardi G.;  
RT "determination of a 15437 bp nucleotide sequence around the inhA gene  
of Mycobacterium avium and similarity analysis of the products of  
putative ORFs.";  
RL Microbiology 144:807-814 (1998).  
CC - SIMILARITY: STRONG, TO M.TUBERCULOSIS RV1480.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AF002133; AAC46198.1; -  
CC KW Hypothetical protein.  
CC SEQUENCE 169 AA; 18686 MW; 37089CA83C906D8E CRC64;

Query Match 97.1%; Score 33; DB 1; Length 169;  
Best Local Similarity 85.7%; Pred. No. 1.4;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRRGNMAY 7  
Db 46 RRRGNMAY 52  
|||||

##### RESULT 2

YB80 MYCTU STANDARD; PRT; 317 AA.  
ID YB80 MYCTU STANDARD; PRT; 317 AA.  
AC P71761; OS3171;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Hypothetical protein RV1480.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.

```

OX NCBI_TaxID=1773;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
EX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrall B.G.;
RA Nature 393:537-544 (1998).
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RL complete genome sequence."
RN [2]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Balcer A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: STRONG, TO M.AVITUM MAV169.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL021184; CAAL6010.1; -
DR EMBL; AE007022; AAK45792.1; ALT_INIT.
DR F1R; C70874; C70874.
DR TIGR; MT1527; -
DR Tuberculist; RV1480; -
DR InterPro; IPR002881; DUF58.
DR InterPro; IPR002035; VWF A.
DR Pfam; PF01882; DUF58; 1.
DR SMART; SMC0327; VWA; 1.
DR Hypothetical protein; Complete proteome.
KW SEQUENCE 317 AA; 34333 MW; 3D1D060FF5ECF56 CRC64;
SQ
Query Match 97.1%; Score 33; DB 1; Length 317;
Best Local Similarity 85.7%; Pred. No. 2.7;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 1 RRRGMAI 7
Db 194 RRRGMV 200
RESULT 3
L2MU_ADE40
ID L2MU_ADE40 STANDARD; PRT; 70 AA.
AC Q64858; Q67717;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Late L2 mu core protein precursor (pMu) (protein X).
EN FX.
OS Human adenovirus type 40.
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
OX NCBI_TaxID=28284;
RN [1]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=Dugan;
EX MEDLINE=85054835; PubMed=6094534;
RA Alestrom P., Akusjaervi G., Lager M., Yeh-Kai L., Pettersson U.;
RT "Genes encoding the core proteins of adenovirus type 2."
RL J. Biol. Chem. 259:13980-13985 (1984).
RN [3]
RX MEDLINE=94087748; PubMed=8263936;
RA Davison A.J., Telford E.A., Watson M.S., McBride K., Mautner V.;
RT "The DNA sequence of adenovirus type 40."
RL J. Mol. Biol. 234:1308-1316 (1993).
RN [2]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=Dugan;
RX MEDLINE=96332521; PubMed=8760431;
RA Grydusuk J.D., Fortsas E., Petric M., Brown M.;
RT "Common epitope on protein VI of enteric adenoviruses from subgenera
RT A and F."
RL J. Gen. Virol. 77:1811-1819 (1996).
CC -1- FUNCTION: THE ROLE OF THE PRECURSOR MIGHT BE TO CONDENSE THE
CC VIRAL PROCHROMATIN FOR ENCAPSIDATION BY VIRTUE OF THE TWO
CC BASIC DOMAINS (BY SIMILARITY).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U14651; AAB19000.1; -
DR EMBL; U14651; AAB19000.1; -
KW Core protein; DNA-binding; Late protein.
FT PROPEP 1 26 BY SIMILARITY.
FT CHAIN 27 40 LATE L2 MU CORE PROTEIN.
FT PROPEP 41 70 BY SIMILARITY.
FT SITE 26 27 CLEAVAGE (BY ADENOVIRUS PROTEASE)
FT SITE 40 41 (POTENTIAL).
FT SITE 31 32 CLEAVAGE (BY ADENOVIRUS PROTEASE)
FT CONFLICT 31 32 RA -> PP (IN REF. 2).
FT SEQUENCE 70 AA; 7596 MW; BF621E015518PF69 CRC64;
SQ
Query Match 88.2%; Score 30; DB 1; Length 70;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 RRRGMA 6
Db 20 RRRGMA 25
RESULT 4
L2MU_ADE02
ID L2MU_ADE02 STANDARD; PRT; 79 AA.
AC F14269;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Late L2 mu core protein precursor (pMu) (11 kDa core protein) (Protein
DE X).
EN FX.
OS Human adenovirus type 2.
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
OX NCBI_TaxID=10515;
RN [1]
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=79211249; PubMed=455453;
RA Zain S., Sambrook J., Roberts R.J., Keller W., Fried M., Dunn A.R.;
RT "Nucleotide sequence analysis of the leader segments in a cloned copy
RT of adenovirus 2 fiber mRNA."
RL Cell 16:851-861 (1979).
RN [2]
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=85054835; PubMed=6094534;
RA Alestrom P., Akusjaervi G., Lager M., Yeh-Kai L., Pettersson U.;
RT "Genes encoding the core proteins of adenovirus type 2."
RL J. Biol. Chem. 259:13980-13985 (1984).
RN [3]

```

PARTIAL SEQUENCE.  
 RX MEDLINE=8818260; PubMed=3357209;  
 RA Weber J.M., Anderson C.W.;  
 RT "Identification of the gene coding for the precursor of adenovirus  
 core protein X";  
 RL J. Virol. 62:1741-1745 (1988).  
 CC -!- FUNCTION: THE ROLE OF THE PRECURSOR MIGHT BE TO CONDENSE THE  
 CC VIRAL PROCHROMATIN FOR ENCAPSIDATION BY VIRTUE OF THE TWO  
 CC BASIC DOMAINS. SUBSEQUENT CLEAVAGE WITHIN THE PARTICLE AFTER  
 CC RESIDUE 31 WOULD RELEASE THE CROSS-LINK AND WOULD PREPARE THE  
 CC VIRAL CHROMATIN FOR THE RELAXED CONFORMATION REQUIRED DURING  
 CC SUBSEQUENT INFECTION AND UNCOATING. THE CLEAVAGE SEEMS TO BE  
 CC NECESSARY FOR INFECTIVITY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; J01317; -; NOT\_ANNOTATED\_CDS.  
 DR PIR; D03837; WMADH2.  
 KW Core protein; DNA-binding; Late protein.  
 FT INIT MET 0  
 FT PROPEP 1 31 LATE L2 MU CORE PROTEIN.  
 FT CHAIN 32 50  
 FT PROPEP 51 79  
 FT PROPEP 31 32 CLEAVAGE (BY ADENOVIRUS PROTEASE).  
 FT SITE 50 51 CLEAVAGE (BY ADENOVIRUS PROTEASE).  
 FT SITE 50 51  
 SQ SEQUENCE 79 AA; 8714 MW; DCD413D3ED0037B CRC64;  
 Query Match 88.2%; Score 30; DB 1; Length 79;  
 Best Local Similarity 100.0%; Pred. No. 3.3;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RRRGWA 6  
 DB 19 RRRGWA 24  
 RESULT 5  
 CTPD MYCTU STANDARD; PRT; 657 AA.  
 AC OS3160;  
 DT 30-MAY-2000 (Rel. 39, Created).  
 DT 30-MAY-2000 (Rel. 39, Last sequence update).  
 DT 15-SEP-2003 (Rel. 42, Last annotation update).  
 DE Probable cation-transporting P-type ATPase D (EC 3.6.3.-).  
 GN CTPD OR RV1469 OR MT1515 OR MTV007.16.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37Rv;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
 RA Badoock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sultun J.E., Taylor K., Whitehead S., Barrell B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 RL complete genome sequence";  
 RL Nature 393:537-544 (1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CDC 1551 / Oshkosh;

RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains";  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 CC -!- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + phosphate.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- SIMILARITY: Belongs to the cation transport ATPases family (P-type  
 CC ATPases). Subfamily IB.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AL021184; CAAL5997.1; -;  
 DR EMBL; AE007021; AAK45780.1; -;  
 DR PIR; H70872; H70872.  
 DR TIGR; MT1515; -;  
 DR Tuberculin; Rv1469;  
 DR InterPro; IPR006416; ATPase-IB\_hvy.  
 DR InterPro; IPR001757; ATPase\_E1-E2.  
 DR InterPro; IPR001366; Cad ATPase.  
 DR InterPro; IPR006404; Heavy met ATPase.  
 DR InterPro; IPR005834; Hydrolase.  
 DR Pfam; PF00122; E1-E2\_ATPase; 1.  
 DR Pfam; PF00702; Hydrolase; 1.  
 DR PRINTS; PR00119; CATATPASE.  
 DR PRINTS; PR00941; CDATPASE.  
 DR TIGRFAMs; TIGR01512; ATPase-IB2\_Cd; 1.  
 DR TIGRFAMs; TIGR01525; ATPase-IB\_hvy; 1.  
 DR TIGRFAMs; TIGR01494; ATPase\_P-Type; 2.  
 DR PROSITE; PS00154; ATPASE\_E1-E2; 1.  
 KW Hydrolase; transmembrane; Phosphorylation; Magnesium; ATP-binding;  
 KW Complete proteome.  
 FT TRANSNEM 40 60 POTENTIAL.  
 FT TRANSNEM 62 82 POTENTIAL.  
 FT TRANSNEM 102 122 POTENTIAL.  
 FT TRANSNEM 288 288 POTENTIAL.  
 FT TRANSNEM 299 319 POTENTIAL.  
 FT TRANSNEM 447 467 POTENTIAL.  
 FT TRANSNEM 477 497 POTENTIAL.  
 FT TRANSNEM 607 627 POTENTIAL.  
 FT MOD RES 347 347 PHOSPHORYLATION (BY SIMILARITY).  
 FT METAL 543 543 MAGNESIUM (BY SIMILARITY).  
 FT METAL 547 547 MAGNESIUM (BY SIMILARITY).  
 SQ SEQUENCE 657 AA; 67884 MW; 1AF5A7DD4BC697D5 CRC64;  
 Query Match 88.2%; Score 30; DB 1; Length 657;  
 Best Local Similarity 85.7%; Pred. No. 28;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RRRGMAI 7  
 DB 399 RRRGMAI 405  
 RESULT 6  
 SPEC\_SCHPO STANDARD; PRT; 716 AA.  
 ID SYEC\_SCHPO  
 AC O13775;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Probable glutamyl-tRNA synthetase, cytoplasmic (EC 6.1.1.17)  
 DE (Glutamyl-tRNA ligase) (Glurs).  
 GN SPAC17A5.15C.

OS Schizosaccharomyces pombe (Fission Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 OX NCBI\_TaxID=4896;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=972;  
 RC MEDLINE=21848401; PubMed=11859360;  
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Brooks K., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Money P., Nourse S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovitch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,  
 RA Welljens I., Vansteels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Wedler L., Wambutt R., Purnelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Ussery D., Barrall B.G., Nurse P.,  
 RA "The genome sequence of Schizosaccharomyces pombe."  
 RL Nature 415:871-880(2002).  
 CC -!- CATALYTIC ACTIVITY: ATP + L-glutamate + tRNA(Glu) = AMP +  
 CC diphosphate + L-glutaminyl-tRNA(Glu).  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; Z98849; CAB11515.1; -  
 DR PIR; T37830; T37830.  
 DR HSP; P00962; 1GTR.  
 DR GeneDB Spombe; SPAC17A5.15c; -  
 DR InterPro; IPR004526; GLTX arch.  
 DR InterPro; IPR000924; Glu tRNA-synt\_1c.  
 DR InterPro; IPR001412; tRNA-synt\_1c.  
 DR Pfam; PF00749; tRNA-synt\_1c; 1.  
 DR Pfam; PF03950; tRNA-synt\_1c; 1.  
 DR PRINTS; PR00987; TRNASYNTHGLJ.  
 DR TIGRFAMs; TIGR00463; GLTX arch; 1.  
 DR PROSITE; PS00178; AA tRNA-LIGASE I; 1.  
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.  
 FT SITE 215 224 "HIGH" REGION.  
 FT SITE 441 445 "KOSKS" REGION.  
 FT BINDING 444 444 ATP (By similarity).  
 FT SEQUENCE 716 AA; 80749 MW; 2ACE0A35ED393227 CRC64;  
 SQ  
 Query Match 88.2%; Score 30; DB 1; Length 716;  
 Best Local Similarity 85.7%; Pred. No. 30;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 RRRGMAI 7  
 DB 471 RRRGMTI 477

RESULT 7  
 XDH\_DROME  
 ID XDH\_DROME STANDARD; PRT; 1335 AA.  
 AC P10351;  
 DT 01-MAR-1989 (Rel. 10, Created)  
 DT 01-MAR-1989 (Rel. 10, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Xanthine dehydrogenase (EC 1.1.1.204) (XDH) (rosy locus protein).  
 GN rosy OR XDH.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE OF 1-231 FROM N.A.  
 RC STRAIN=Canton-S;  
 RC MEDLINE=87248039; PubMed=3036645;  
 RA Lee C.S., Curtis D., Gray M., Bender W.;  
 RT "Mutations affecting expression of the rosy locus in Drosophila  
 RT melanogaster.";  
 RL Genetics 116:55-66(1987).  
 RN [2]  
 RP SEQUENCE OF 199-1335 FROM N.A.  
 RC STRAIN=Canton-S;  
 RC MEDLINE=87248040; PubMed=3036646;  
 RA Keith T.P., Riley M.A., Kreitman M., Lewontin R.C., Curtis D.,  
 RA Chambers G.;  
 RT "Sequence of the structural gene for xanthine dehydrogenase (rosy  
 RT locus) in Drosophila melanogaster.";  
 RL Genetics 116:67-73(1987).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Canton-S;  
 RA Riley M.;  
 RL Submitted (FEB-1987) to the EMBL/GenBank/DBJ databases.  
 CC -!- CATALYTIC ACTIVITY: Xanthine + NAD(+) + H(2)O = urate + NADH.  
 CC -!- COFACTOR: FAD, MOLYBDOPTERIN, AND TWO 2FE-2S CLUSTERS.  
 CC -!- SUBUNIT: Homodimer.  
 CC -!- SUBCELLULAR LOCATION: Peroxisomal.  
 CC -!- SIMILARITY: BELONGS TO THE XANTHINE DEHYDROGENASE FAMILY.  
 CC -!- SIMILARITY: TO 2FE-2S FERREDOXINS IN THE N-TERMINAL DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; Y00308; CAA68409.1; -  
 DR PIR; S07245; S07245.  
 DR HSP; P80457; 1FO4.  
 DR FlyBase; FBgn0003308; ry.  
 DR InterPro; IPR002888; 2Fe-2S bind.  
 DR InterPro; IPR006056; 2Fe2S-ferredoxin.  
 DR InterPro; IPR000674; Aldxan\_dh\_C.  
 DR InterPro; IPR005107; CO deh flav\_C.  
 DR InterPro; IPR002346; dehydrog\_molyb.  
 DR InterPro; IPR000572; Euk Mb oxred.  
 DR InterPro; IPR001041; Ferredoxin.  
 DR Pfam; PF02738; Ald\_Xan\_dh\_C2; 1.  
 DR Pfam; PF01315; Ald\_Xan\_dh\_C; 1.  
 DR Pfam; PF03450; CO deh flav\_C; 1.  
 DR Pfam; PF00941; FAD binding\_5; 1.  
 DR Pfam; PF00111; fer2; 1.  
 DR Pfam; PF01799; fer2\_2; 1.  
 DR ProDom; PD186071; 2Fe-2S bind; 1.  
 DR PROSITE; PS00197; 2FE2S FERREDOXIN; 1.  
 DR PROSITE; PS00559; MOLYBDOPTERIN\_EUK; 1.



KW Oxidoreductase; NAD; Molybdenum; Flavoprotein; FAD; Metal-binding;  
 KW Iron-sulfur; Iron; 2Fe-2S; Peroxisome.  
 FT METAL 37 37 IRON-SULFUR (2FE-2S) (BY SIMILARITY).  
 FT METAL 43 43 IRON-SULFUR (2FE-2S) (BY SIMILARITY).  
 FT METAL 48 48 IRON-SULFUR (2FE-2S) (BY SIMILARITY).  
 FT METAL 51 51 IRON-SULFUR (2FE-2S) (BY SIMILARITY).  
 SQ SEQUENCE 1335 AA; 146925 MW; B37C5F439303589 CRC64;

Query Match 88.2%; Score 30; DB 1; Length 1335;  
 Best Local Similarity 71.4%; Pred. No. 57;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RREGMAI 7  
 | : | | | :  
 Db 999 RKGMAV 1005

## RESULT 8

ID\_XDH\_DROPS STANDARD; PRT; 1342 AA.  
 AC P22811;  
 DT 01-AUG-1991 (Rel. 19, Created)  
 DT 01-AUG-1991 (Rel. 19, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Xanthine dehydrogenase (SC 1.1.1.204) (XDH) (Rosy locus protein).  
 GN RY OR XDH.  
 OS Drosophila pseudoobscura (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7237;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=69158785; PubMed=2493563;  
 RA Ryley M.A.;  
 RT "Nucleotide sequence of the Xdh region in Drosophila pseudoobscura  
 and an analysis of the evolution of synonymous codons.";  
 RL Mol. Biol. Evol. 6:33-52(1989).  
 CC -1- CATALYTIC ACTIVITY: Xanthine + NAD(+) + H(2)O = urate + NADH.  
 CC -1- COFACTOR: FAD, MOLYBDOPTERIN, AND TWO 2FE-2S CLUSTERS.  
 CC -1- SUBUNIT: Homodimer.  
 CC -1- SUBCELLULAR LOCATION: Peroxisomal.  
 CC -1- SIMILARITY: BELONGS TO THE XANTHINE DEHYDROGENASE FAMILY.  
 CC -1- SIMILARITY: TO 2FE-2S FERREDOXINS IN THE N-TERMINAL DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; M33977; AAA29022.1; -  
 DR PIR; A31946; A31946.  
 DR HSP; P80457; IFO4.  
 DR FlyBase; F5gn0012736; Dpsee\ry.  
 DR InterPro; IPR002888; 2fe-2s\_bind.  
 DR InterPro; IPR006058; 2fe2s\_ferredoxin.  
 DR InterPro; IPR000674; Aldxan\_dh\_C.  
 DR InterPro; IPR005107; CO\_deh\_flav\_C.  
 DR InterPro; IPR002346; dehydrog\_molyb.  
 DR InterPro; IPR000572; Fek\_Mb\_Cxrd.  
 DR InterPro; IPR01041; Ferredoxin.  
 DR Pfam; PF02738; Ald\_Xan\_dh\_C2; 1.  
 DR Pfam; PF01315; Ald\_Xan\_dh\_C; 1.  
 DR Pfam; PF03450; CO\_deh\_flav\_C; 1.  
 DR Pfam; PF00941; FAD\_binding\_5; 1.  
 DR Pfam; PF00111; fer2; 1.  
 DR Pfam; PF01799; fer2; 1.  
 DR ProDom; PD186071; 2Fe-2S\_bind; 1.  
 DR ProSITE; PS00197; 2FE2S\_FERREDOXIN; 1.  
 DR ProSITE; PS00559; MOLYBDOPTERIN\_EUK; 1.

KW Oxidoreductase; NAD; Molybdenum; Flavoprotein; FAD; Metal-binding;  
 KW Iron-sulfur; Iron; 2Fe-2S; Peroxisome.  
 FT METAL 41 41 IRON-SULFUR (2FE-2S) (BY SIMILARITY).  
 FT METAL 47 47 IRON-SULFUR (2FE-2S) (BY SIMILARITY).  
 FT METAL 52 52 IRON-SULFUR (2FE-2S) (BY SIMILARITY).  
 FT METAL 55 55 IRON-SULFUR (2FE-2S) (BY SIMILARITY).  
 SQ SEQUENCE 1342 AA; 147422 MW; 169254E4FAAE021 CRC64;

Query Match 88.2%; Score 30; DB 1; Length 1342;  
 Best Local Similarity 71.4%; Pred. No. 58;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RREGMAI 7  
 | : | | | :  
 Db 1007 RKGMAV 1013

## RESULT 9

ID\_YHH6\_YEAST STANDARD; PRT; 382 AA.  
 AC P32793;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE Hypothetical 41.8 kDa protein in SPO13-ARG4 intergenic region.  
 GN YHR016C OR YSC84.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE OF 87-382 FROM N.A.  
 RX MEDLINE=94078673; PubMed=8256520;  
 RA Rocco V.; Daly M.J.; Matre V.; Lichten M.; Nicolas A.;  
 RT "Identification of two divergently transcribed genes  
 centromere-proximal to the ARG4 locus on chromosome VIII of  
 Saccharomyces cerevisiae.";  
 RL Yeast 9:1111-1120(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288c / AB972;  
 RX MEDLINE=94378003; PubMed=8091229;  
 RA Johnston M.; Andrews S.; Brinkman R.; Cooper J.; Ding H.; Dover J.;  
 Du Z.; Favell A.; Fulton L.; Gattung S.; Geisel C.; Kirsten J.;  
 RA Kucaba T.; Hillier L.; Jier M.; Johnston L.; Langston Y.;  
 RA Latreille P.; Louis E.J.; Macri C.; Mardis E.; Manes S.; Mouser L.;  
 Rhan M.; Rifkin L.; Riles L.; St Peter H.; Trevaskis E.; Vaughan K.;  
 Vignati D.; Wilcox L.; Wohlman P.; Waterston R.; Wilson R.;  
 Vaudin M.;  
 RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome  
 VIII.";  
 RL Science 265:2077-2082(1994).

CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
 CC -1- SIMILARITY: STRONG, TO YEAST YFR024C.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; U10400; AAB68945.1; ALT\_INIT.  
 DR EMBL; L06795; AAA56990.1; ALT\_INIT.  
 DR PIR; S46791; S46791.  
 DR HSP; P29355; ISEM.  
 DR SGD; S0001058; YSC84.  
 DR GO; GO:0005857; C:actin cortical patch (sensu Saccharomyces); IDA.  
 DR GO; GO:0007015; P:actin filament organization; IPI.  
 DR GO; GO:0006897; P:endocytosis; IMP.  
 DR InterPro; IPR001452; SH3.  
 DR Pfam; PF04366; DUF500; 1.

```

KW Pfam; PF00018; SH3; 1.
DR ProDom; PD000066; SH3; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00002; SH3; 1.
KW Hypothetical protein; SH3 domain.
FT DOMAIN 323 382 SH3.
SQ SEQUENCE 382 AA; 41829 MW; E3959F3745C0365D CRC64;

Query Match      85.3%; Score 29; DB 1; Length 382;
Best Local Similarity 71.4%; Pred. No. 28;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRRGMAI 7
Db 183 RRRGMSL 189

RESULT 10
SYEC YEAST STANDARD; PRT; 724 AA.
AC P46655;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE GlutamyI-tRNA synthetase, cytoplasmic (EC 6.1.1.17) (Glutamate-tRNA
DE ligase) (GluRS) (P85).
GN YGL245W OR G0583 OR HRB724.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OC NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Frantz J.D., Gilbert W.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Vandenbol M., Durand P., Portetelle D., Hilger F.;
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-146 FROM N.A.
RC STRAIN=S288c / FV1679;
RA Coissac E., Maillier E., Netter P.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: ATP + L-glutamate + tRNA(Glu) = AMP +
CC diphosphate + L-glutamyl-tRNA(Glu).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U32265; AAA78905.1; -
DR EMBL; Z49149; CAAB9009.1; -
DR EMBL; Z72767; CAA96984.1; -
DR EMBL; X94357; CAA64142.1; -
DR PIR; S53934; S53934.
DR HSSP; P00962; 1GTR.
DR SGD; S0003214; YGL245W.
DR InterPro; IPR004526; Glx arch.
DR InterPro; IPR000924; Glu tRNA-synt_1c.
DR InterPro; IPR001412; tRNA-synt_1.
DR Pfam; PF00749; tRNA-synt_1c; 1.
DR Pfam; PF03950; tRNA-synt_1c.C; 1.
DR PRINTS; PR00987; TRNASYNTHGLU.
DR TIGRFAMs; TIGR00463; gltx arch; 1.
DR PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.

```

Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.  
"HIGH" REGION.  
"KMSKS" REGION.  
SITE 226 235  
SITE 453 457  
FT BINDING 456 456  
E -> D (IN REF. 1).  
FT CONFLICT 225 225  
V -> A (IN REF. 1).  
FT CONFLICT 489 489  
P -> S (IN REF. 1).  
FT CONFLICT 526 526  
P -> M (IN REF. 1).  
FT CONFLICT 562 562  
FT CONFLICT 714 724  
GKSNVYKARK -> VNLSSTMYQVKNKHHSNVTYLCYFS  
TSTF (IN REF. 1).  
SQ SEQUENCE 724 AA; 82662 MW; 34669BFB69CD41BE CRC64;

Query Match 85.3%; Score 29; DB 1; Length 724;  
Best Local Similarity 71.4%; Pred. No. 53;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RRRGMAI 7  
Db 483 RRRGMTV 489

RESULT 11  
ICE6 HUMAN  
ID ICE6 HUMAN STANDARD; PRT; 293 AA.  
AC P55312; Q9BQ87;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Caspase-6 precursor (EC 3.4.22.-) (Apoptotic protease Mch-2).  
GN CASP6 OR MCH2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OC NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM ALPHA AND BETA).  
RC TISSUE=Lung;  
RX MEDLINE=95316841; PubMed=7796396;  
RA Fernandes-Alnemri T., Litwack G., Alnemri E.S.;  
RT "Mch2, a new member of the apoptotic Ced-3/Ice cysteine protease gene  
RT family";  
RL Cancer Res. 55:2737-2742(1995).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM ALPHA).  
RC TISSUE=Lung;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Frange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Munz D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahney J., Heiton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalios D.E.,  
RA Schnarch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [3]  
RP PROCESSING.  
RC TISSUE=Lymphocytes;  
RX MEDLINE=97059171; PubMed=8900201;  
RA Srinivasula S.M., Fernandes-Alnemri T., Zangrilli J., Robertson N.,  
RA Armstrong R.C., Wang L., Trapani J.A., Tomaselli K.J., Litwack G.,  
RA Alnemri E.S.;

RT The Ced-3/interleukin beta converting enzyme-like homolog Mch6 and  
the lamin-cleaving enzyme Mch2alpha are substrates for the apoptotic  
mediator CPP32.";  
RL J. Biol. Chem. 271:27099-27106(1996).  
CC -1- FUNCTION: INVOLVED IN THE ACTIVATION CASCADE OF CASPASES  
RESPONSIBLE FOR APOPTOSIS EXECUTION. CLEAVES POLY(ADP-RIBOSE)  
CC POLYMERASE IN VITRO, AS WELL AS LAMINS. OVEREXPRESSION PROMOTES  
CC PROGRAMMED CELL DEATH.  
CC -1- SUBUNIT: HETERODIMER OF A 18 kDa (P18) AND A 11 kDa (P11) SUBUNIT.  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -1- ALTERNATIVE PRODUCTS:  
CC Event-Alternative splicing; Named isoforms=2;  
CC Name=Alpha;  
CC IsoId=P55212-1; Sequence=Displayed;  
CC Name=Beta; IsoId=P5212-2; Sequence=VSP\_000805;  
CC IsoId=P5212-2; Sequence=VSP\_000805;  
CC -1- PTM: CLEAVAGES BY CPP32, CASPASE-8 OR -10 GENERATE THE TWO ACTIVE  
CC SUBUNITS.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C14.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U20536; AAC50168.1; -;  
DR EMBL; U20537; AAC50169.1; -;  
DR EMBL; BC000305; AAH00305.1; -;  
DR EMBL; BC004460; AAH04460.1; -;  
DR HSP; P42574; IPAU.  
DR MEROPS; C14.005; -;  
DR Genew; HGNC:1507; CASP6.  
DR MIM; 601532; -;  
DR GO; GO:0008234; F:cysteine-type peptidase activity; TAS.  
DR GO; GO:0006917; P:induction of apoptosis; TAS.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; TAS.  
DR InterPro; IPR002398; ICE.  
DR InterPro; IPR002138; ICE\_P10.  
DR InterPro; IPR001309; ICE\_P20.  
DR Pfam; PF00655; ICE\_P10; 1.  
DR Pfam; PF00656; ICE\_P20; 1.  
DR PRINTS; PR00376; ILIBENZTYME.  
DR SMART; SM00115; CASG; 1.  
DR PROSITE; PS01122; CASPASE\_CYS; 1.  
DR PROSITE; PS01121; CASPASE\_HIS; 1.  
DR PROSITE; PS0207; CASPASE\_P10; 1.  
DR PROSITE; PS0208; CASPASE\_P20; 1.  
KW Hydroxylase, Thiol protease, Apoptosis; Zymogen; Alternative splicing.  
FT PROPEP 1 23  
FT CHAIN 24 179 CASPASE-6 SUBUNIT P18.  
FT PROPEP 180 193  
FT CHAIN 194 293 CASPASE-6 SUBUNIT P11.  
FT ACT\_SITE 121 121 BY SIMILARITY.  
FT ACT\_SITE 163 163 BY SIMILARITY.  
FT VARSPIC 14 102 Missing (in isoform Beta).  
FT CONFLICT 66 66 /FTID=VSP\_000805.  
FT CONFLICT 66 66 G -> R (IN REF. 1).  
SQ SEQUENCE 293 AA; 33310 MW; 0738AB4F971EBD7 CRC64;  
  
Query Match 82.4%; Score 28; DB 1; Length 293;  
Best Local Similarity 71.4%; Pred. No. 36;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 RRRGMAI 7  
DB 42 RRRGIAL 48  
  
RESULT 12  
Y812\_ARCFU

Y812\_ARCFU STANDARD; PRT; 298 AA.  
AC 029446; 2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein AF0812.  
GN AF0812.  
OS Archaeoglobus fulgidus.  
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;  
OC Archaeoglobaceae; Archaeoglobus.  
OX NCBI\_TaxID=2234;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;  
RX MEDLINE=98049343; PubMed=9389475;  
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,  
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,  
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,  
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,  
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,  
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,  
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,  
RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,  
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,  
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,  
RA Venter J.C.;  
RT "The complete genome sequence of the hyperthermophilic, sulphate-  
reducing archaeon Archaeoglobus fulgidus.";  
RL Nature 390:364-370(1997).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AE001048; AAB90432.1; -;  
DR FIC; D69351; D69351.  
DR TIGR; AF0812; -;  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 298 AA; 34385 MW; ED59B86A07AC5A30 CRC64;  
  
Query Match 82.4%; Score 28; DB 1; Length 298;  
Best Local Similarity 71.4%; Pred. No. 37;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 RRRGMAI 7  
DB 45 RRRGMVEV 51  
  
RESULT 13  
YHCH\_BACSU STANDARD; PRT; 305 AA.  
ID YHCH\_BACSU  
AC P54592;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Hypothetical ABC transporter ATP-binding protein yhch.  
GN YHCH.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=97124185; PubMed=8959498;  
RA Noback M.A., Terpestra P., Holsappel S., Venema G., Bron S.;  
RT "A 22 kb DNA sequence in the cspB-glpPFD region at 75 degrees on the  
Bacillus subtilis chromosome.";  
RL Microbiology 142:3021-3026(1996).

RN SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=98044033; PubMed=9384377;  
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
 RA Azevedo V., Bartero M.G., Bessieres P., Bolotin A., Borchert S.,  
 RA Brattis R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
 RA Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,  
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.I.,  
 RA Ettian K.D., Errington J., Fabret C., Ferrai E., Fougier D.,  
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,  
 RA Gims S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,  
 RA Guiseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,  
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
 RA Joris B., Karamata D., Kasahara Y., Klaer-Blanchard M., Klein C.,  
 RA Kobayashi Y., Koetter P., Konigstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
 RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,  
 RA Sorokin A., Taconi E., Takagi T., Takahashi H., Tognoni K.,  
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni K.,  
 RA Terasaki V., Uchiyama S., Vandenberg M., Vannier F., Vassarotti A.,  
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,  
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,  
 RA "The complete genome sequence of the Gram-positive bacterium Bacillus  
 subtilis".  
 RL Nature 390:249-256(1997).  
 CC -!- SIMILARITY: Belongs to the ABC transporter family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X96983; CAA65691.1; -;  
 DR EMBL; Z99108; CAB12736.1; -;  
 DR PIR; D69822; D69822.  
 DR Subtilist; BG11586; YhCh.  
 DR InterPro; IPR003593; AAA ATPase.  
 DR InterPro; IPR003439; ABC\_transporter.  
 DR Pfam; PF00005; ABC\_tran; 1.  
 DR ProDom; PD000006; ABC\_transporter; 1.  
 DR SMART; SM00382; AAA; 1.  
 DR PROSITE; PS00211; ABC\_TRANSPORTER\_1; 1.  
 DR PROSITE; PS00893; ABC\_TRANSPORTER\_2; 1.  
 KW Hypothetical protein; ATP-binding; Transport; Complete proteome.  
 FT NP\_BIND 37 44  
 FT ATP (POTENTIAL).  
 SQ SEQUENCE 305 AA; 34470 MW; 1EBD2DC4.10BFFB7F CRC64;  
 Query Match 82.4%; Score 28; DB 1; Length 305;  
 Best Local Similarity 71.4%; Pred. No. 38;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 RERGVAI 7  
 DB 180 RERGVAV 186  
 RESULT 14  
 ID\_HUMAN  
 ID\_ADA\_HUMAN  
 AC P00813;

DT 21-JUL-1986 (Rel. 01, Created)  
 DT 23-OCT-1986 (Rel. 02, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Adenosine deaminase (EC 3.5.4.4) (Adenosine aminohydrolase).  
 GN ABA.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=85257473; PubMed=3839456;  
 RA Valerio D., Duyvesteyn M.G.C., Dekker B.M.M., Weeda G.,  
 RA Berkvens T.M., van der Voort L., van Ormondt H., van der Eb A.J.,  
 RA "Adenosine deaminase: characterization and expression of a gene with  
 RT a remarkable promoter".  
 RL EMBO J. 4:437-443(1985).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87128922; PubMed=3028473;  
 RA Wiginton D.A., Kaplan D.J., States J.C., Akesson A.L., Perme C.M.,  
 RA Bilyk I.J., Vaughn A.J., Lattier D.B., Hutton J.J.,  
 RA "Complete sequence and structure of the gene for human adenosine  
 RT deaminase".  
 RL Biochemistry 25:8234-8244(1986).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=85006946; PubMed=6090454;  
 RA Daddona P.E., Shewach D.S., Kelley W.N., Argos P., Markham A.F.,  
 RA Orkin S.H.,  
 RA "Human adenosine deaminase. cDNA and complete primary amino acid  
 RT sequence".  
 RL J. Biol. Chem. 259:12101-12106(1984).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=84169545; PubMed=6546794;  
 RA Wiginton D.A., Adrian G.S., Hutton J.J.,  
 RA "Sequence of human adenosine deaminase cDNA including the coding  
 RT region and a small intron".  
 RL Nucleic Acids Res. 12:2439-2446(1984).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.L., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan B., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Rulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchan J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield A.S., Krzywinski M.I., Skalska U., Marra M.A.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,  
 RA "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences".  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [6]  
 RP SEQUENCE OF 141-363 FROM N.A.  
 RX MEDLINE=84032477; PubMed=668808;  
 RA Orkin S.H., Daddona P.E., Shewach D.S., Markham A.F., Bruns G.A.,  
 RA Goff S.C., Kelley W.N.,  
 RA "Molecular cloning of human adenosine deaminase gene sequences".  
 RL J. Biol. Chem. 258:12753-12756(1983).  
 RN [7]  
 RP VARIANT ADA\*2 ASN-8.

RX MEDLINE=94304087; PubMed=8031011;  
 RA Hirschhorn R., Yang D.R., Ierani A.;  
 RA "An Aspaan substitution results in the adenosine deaminase (ADA)  
 RT genetic polymorphism (ADA 2 allele); occurrence on different  
 RT chromosomal backgrounds and apparent intragenic crossover.";  
 RL Ann. Hum. Genet. 58:1-9 (1994).  
 RN [9]  
 RN VARIANTS SCID.  
 RP MEDLINE=85036320; PubMed=6208479;  
 RX MEDLINE=85036320; PubMed=6208479;  
 RA Adrian G.S., Wiginton D.A., Hutton J.;  
 RA "Structure of adenosine deaminase mRNAs from normal and adenosine  
 RT deaminase-deficient human cell lines.";  
 RL Mol. Cell. Biol. 4:1712-1717 (1984).  
 RN [9]  
 RN VARIANTS SCID GLN-101.  
 RP MEDLINE=85289940; PubMed=3839802;  
 RX MEDLINE=85289940; PubMed=3839802;  
 RA Bonthron D.T., Markham A.F., Ginsburg D., Orkin S.H.;  
 RA "Identification of a point mutation in the adenosine deaminase gene  
 RT responsible for immunodeficiency.";  
 RL J. Clin. Invest. 76:894-897 (1985).  
 RN [10]  
 RN VARIANTS SCID TRP-101; HIS-211 AND VAL-329.  
 RP MEDLINE=89034097; PubMed=3182793;  
 RX MEDLINE=89034097; PubMed=3182793;  
 RA Akesson A.L., Wiginton D.A., Dusing M.R., States J.C., Hutton J.J.;  
 RA "Mutant human adenosine deaminase alleles and their expression by  
 RT transfection into fibroblasts.";  
 RL J. Biol. Chem. 263:16291-16296 (1988).  
 RN [11]  
 RN VARIANTS SCID GLN-297.  
 RP MEDLINE=89109562; PubMed=2783588;  
 RX MEDLINE=89109562; PubMed=2783588;  
 RA Hirschhorn R., Tzall S., Ellenbogen A., Orkin S.H.;  
 RA "Identification of a point mutation resulting in a heat-labile  
 RT adenosine deaminase (ADA) in two unrelated children with partial ADA  
 RT deficiency.";  
 RL J. Clin. Invest. 83:497-501 (1989).  
 RN [12]  
 RN VARIANTS SCID CYS-156 AND LEU-291.  
 RP MEDLINE=93244849; PubMed=1284479;  
 RX MEDLINE=93244849; PubMed=1284479;  
 RA Hirschhorn R.;  
 RA "Identification of two new missense mutations (R156C and S291L) in  
 RT two ADA-SCID patients unusual for response to therapy with partial  
 RT exchange transfusions.";  
 RL Hum. Mutat. 1:166-168 (1992).  
 RN [13]  
 RN VARIANTS SCID LEU-101; HIS-156 AND MET-177.  
 RP MEDLINE=94043746; PubMed=8227344;  
 RX MEDLINE=94043746; PubMed=8227344;  
 RA Santisteban I., Arredondo-Vega F.X., Kelly S., Mary A., Fischer A.,  
 RA Hummel D.S., Lawton A., Sorensen R.U., Stiehm E.R., Uribe L.,  
 RA Weinberg K., Hershfield M.S.;  
 RA "Novel splicing, missense, and deletion mutations in seven adenosine  
 RT deaminase-deficient patients with late/delayed onset of combined  
 RT immunodeficiency disease. Contribution of genotype to phenotype.";  
 RL J. Clin. Invest. 92:2291-2302 (1993).  
 RN [14]  
 RN VARIANT SCID ARG-20.  
 RP MEDLINE=94130402; PubMed=8299233;  
 RX MEDLINE=94130402; PubMed=8299233;  
 RA Yang D.R., Huie M.L., Hirschhorn R.;  
 RA "Homozygosity for a missense mutation (G20R) associated with neonatal  
 RT onset adenosine deaminase-deficient severe combined immunodeficiency  
 RT (ADA-SCID).";  
 RL Clin. Immunol. Immunopathol. 70:171-175 (1994).  
 RN [15]  
 RN VARIANT SCID GLN-142 AND VARIANT ARG-80.  
 RP MEDLINE=96154678; PubMed=8589684;  
 RX MEDLINE=96154678; PubMed=8589684;  
 RA Santisteban I., Arredondo-Vega F.X., Kelly S., Loubser M., Meydan N.,  
 RA Roifman C., Howell P.L., Bowen T., Weinberg K.I., Schroeder M.L.,  
 RA Hershfield M.S.;  
 RA "Three new adenosine deaminase mutations that define a splicing  
 RT enhancer and cause severe and partial phenotypes: implications for  
 RT evolution of a Cpg hotspot and expression of a transduced ADA cDNA.";  
 RL Hum. Mol. Genet. 4:2081-2087 (1995).  
 RN [16]  
 RN VARIANTS SCID ASP-15; ASP-83 AND ASP-179.

RX MEDLINE=95322982; PubMed=7599635;  
 RA Santisteban I., Arredondo-Vega F.X., Kelly S., Debre M., Fisher A.,  
 RA Perignon J.L., Hilman B., Eldahr J., Dreyfus D.H., Gelfand E.W.,  
 RA Howell P.L., Hershfield M.S.;  
 RA "Four new adenosine deaminase mutations, altering a zinc-binding  
 RT histidine, two conserved alanines, and a 5' splice site.";  
 RL Hum. Mutat. 5:243-250 (1995).  
 RN [17]  
 RN VARIANTS SCID MET-152 AND ILE-233.  
 RP MEDLINE=97369476; PubMed=9225964;  
 RX MEDLINE=97369476; PubMed=9225964;  
 RA Hirschhorn R., Borkowsky W., Jiang C.-K., Yang D.R., Jenkins T.;  
 RA "Two newly identified mutations (Thr233Ile and Leu152Met) in  
 RT partially adenosine deaminase-deficient (ADA-) individuals that  
 RT result in differing biochemical and metabolic phenotypes.";  
 RL Hum. Genet. 100:22-29 (1997).  
 RN [18]  
 RN VARIANTS SCID CYS-74; MET-129; GLU-140; TRP-149 AND PRO-199.  
 RP MEDLINE=97369476; PubMed=9225964;  
 RX MEDLINE=97369476; PubMed=9225964;  
 RA Arredondo-Vega F.X., Santisteban I., Notarangelo L.D., el Dahr J.,  
 RA Buckley R., Roifman C., Conley M.E., Hershfield M.S.;  
 RA "Seven novel mutations in the adenosine deaminase (ADA) gene in  
 RT patients with severe and delayed onset combined immunodeficiency:  
 RT G74C, V129M, G140E, R149W, Q199P, 462delG, and E337del.";  
 RL Hum. Mutat. 11:482-482 (1998).  
 CC -!- CATALYTIC ACTIVITY: Adenosine + H(2)O = inosine + NH(3).  
 CC -!- TISSUE SPECIFICITY: FOUND IN ALL TISSUES, OCCURS IN LARGE AMOUNTS  
 CC IN T-LYMPHOCYTES AND, AT THE TIME OF WEANING, IN GASTROINTESTINAL  
 CC TISSUES.  
 CC -!- POLYMORPHISM: THERE IS A COMMON ALLELE, ADA\*2, ALSO KNOWN AS THE  
 CC ADA 2 ALLOZYME.  
 CC -!- DISEASE: Defects in ADA are a cause of autosomal recessive severe  
 CC combined immuno-deficiency (SCID) [MIM:102700]. SCID is a  
 CC congenital disorder characterized by impairment of both humoral  
 CC and cell-mediated immunity, leukopenia, and low or absent antibody  
 CC levels. Onset is during infancy. Less often, immune dysfunction  
 CC develops later in childhood (delayed) and in a few cases ADA  
 CC deficiency has been diagnosed in chronically ill teenagers and  
 CC adults (late or adult onset). Population and newborn screening  
 CC programs have also identified several healthy individuals with  
 CC normal immunity, mainly in african descent, who have partial ADA  
 CC deficiency. Most patients are compound heterozygotes. SCID arises  
 CC if both mutations eliminate ADA activity, whereas a single allele  
 CC with residual activity can confer a milder phenotype. ADA  
 CC deficiency accounts for about one-half of cases of autosomal  
 CC recessive SCID.  
 CC -!- DISEASE: In hereditary hemolytic anemia, the level of this enzyme  
 CC in erythrocytes increases 50-70 times.  
 CC -!- SIMILARITY: Belongs to the adenosine and AMP deaminases family.  
 CC -!- DATABASE: NAME=ADABase; NOTE=ADA mutation db;  
 CC WWW="http://www.uta.fi/mt/bioinfo/ADABase/".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

Query Match 82.4%; Score 28; DB 1; Length 363;  
 Best Local Similarity 71.4%; Pred. No. 45;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRRGMAI 7  
 DB 32 RRRGIAL 38

RESULT 15  
 COBB RHIME  
 ID COBB RHIME STANDARD; PRT; 429 AA.  
 AC Q32P18;  
 DT 15-SEP-2003 (Rel. 42, Created)  
 DT 15-SEP-2003 (Rel. 42, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Cobytrinic acid A,C-diamide synthase.  
 GN COBB OR R01945 OR SMC04282.  
 OS Rhizobium meliloti (Sinorhizobium meliloti).  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;

OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.  
OX NCBI\_TaxID=382;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1021;  
RX MEDLINE=21396507; PubMed=11481430;  
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,  
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,  
RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,  
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,  
RA Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.;  
RT "Analysis of the chromosome sequence of the legume symbiont  
RT Sinorhizobium meliloti strain 1021.";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).  
CC -!- FUNCTION: Responsible for the amidation of carboxylic groups at  
CC position A and C of either cobytrinic acid or hydrogenobrynic acid.  
CC NH(2) groups are provided by glutamine, and one molecule of ATP is  
CC hydrogenolyzed for each amidation (By similarity).  
CC -!- PATHWAY: Cobalamin biosynthesis.  
CC -!- SIMILARITY: Belongs to the cobS/cobQ family. CobS subfamily.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AL591789; CAC46524.1; -.  
DR HAMAP; MF 00027; -; 1.  
DR InterPro; IPR004484; CbiA.  
DR InterPro; IPR002586; CbiA\_P.  
DR Pfam; PF01656; CbiA; 1.  
DR TIGRFAMs; TIGR00379; cobS; 1.  
KW Cobalamin biosynthesis; Complete proteome.  
SQ SEQUENCE 429 AA; 45086 MW; E6D5670CC4B4FF17 CRC64;  
  
Query Match 82.4%; Score 28; DB 1; Length 429;  
Best Local Similarity 71.4%; Pred. No. 53;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 RRRGMAI 7  
Db 27 RRRGLSI 33  
  
Search completed: February 11, 2004, 17:04:08  
Job time : 7.16867 secs

GenCore version 5.1.6  
Copyright (C) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 16:55:29 ; Search time 9.33333 Seconds  
(without alignments)  
72.127 Million cell updates/sec

Title: US-09-901-187C-4

Perfect score: 34

Sequence: 1 RRRGMWAI 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: Pir1.\*
- 2: Pir2.\*
- 3: Pir3.\*
- 4: Pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Query Match %	Score	Length	DB	ID	Description
1	33	97.1	317	2	C70874	hypothetical prote
2	33	97.1	320	2	B87135	conserved hypotet
3	32	94.1	267	2	T37047	probable nitrate r
4	30	88.2	80	1	WVADH2	late L2 mu core pr
5	30	88.2	185	2	T08768	hypothetical prote
6	30	88.2	263	2	A87149	conserved hypotet
7	30	88.2	321	2	H83560	probable type II s
8	30	88.2	567	2	T16625	asparagine synthas
9	30	88.2	657	2	H70872	probable ctpD prot
10	30	88.2	716	2	T37830	probable glutamate
11	30	88.2	943	2	T34847	probable transcrip
12	30	88.2	1335	2	S07245	xanthine dehydroge
13	30	88.2	1342	2	A31946	xanthine dehydroge
14	29	85.3	62	2	A93805	hypothetical prote
15	29	85.3	62	2	A29278	hypothetical prote
16	29	85.3	171	2	D95336	hypothetical prote
17	29	85.3	246	2	A25304	hypothetical prote
18	29	85.3	264	2	D70804	hypothetical prote
19	29	85.3	298	2	C75444	hypothetical prote
20	29	85.3	359	2	D83103	probable phospholi
21	29	85.3	361	2	B75444	probable phosphino
22	29	85.3	468	2	S46791	hypothetical prote
23	29	85.3	724	1	S63934	glutamate-tRNA lig
24	28	82.4	60	2	A62568	hypothetical prote
25	28	82.4	298	2	D69351	hypothetical prote
26	28	82.4	305	2	D69822	ABC transporter (A
27	28	82.4	316	2	A57356	[acyl-carrier-prot
28	28	82.4	356	2	T33984	hypothetical prote
29	28	82.4	363	1	DUHUA	adenosine deaminas

30	28	82.4	375	2	T35015	probable 3-oxoadip
31	28	82.4	450	2	D87342	multidrug resistan
32	28	82.4	617	2	JC5721	vacuolar protein s
33	28	82.4	667	2	F70682	probable membrane
34	28	82.4	703	2	S08119	heat shock protein
35	28	82.4	703	2	A44983	heat shock protein
36	28	82.4	908	2	T07409	lipoygenase (BC 1
37	28	82.4	914	2	T07065	probable lipoxigen
38	28	82.4	926	2	E96749	probable lipoxigen
39	28	82.4	936	2	T06190	lipoxigenase (EC 1
40	28	82.4	1065	2	A43421	SEC8 protein - yea
41	28	82.4	1364	2	T51920	probable xanthine
42	28	82.4	1461	2	B70588	probable polyketid
43	27	79.4	75	2	AD3131	hypothetical prote
44	27	79.4	93	2	S34799	nodulin (clone GMN
45	27	79.4	99	2	C70694	hypothetical prote

## ALIGNMENTS

### RESULT 1

C70874

hypothetical protein Rv1480 - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999

C:Accession: C70874

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, J.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S. Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: C70874

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-317 <COL>

A:Cross-references: GB:AL021184; GB:AL123456; NID:G3261498; PIDN:CAAL6010.1; PID:el2997

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: Rv1480

Query Match 97.1%; Score 33; DB 2; Length 317;  
Best Local Similarity 85.7%; Pred. No. 11;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRRGMWAI 7

Db 194 RRRGMWAV 200

### RESULT 2

B87135

conserved hypothetical protein ML1809 [imported] - Mycobacterium leprae

C:Species: Mycobacterium leprae

C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001

C:Accession: B87135

R:Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; F. R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd eam, M.A.; Rutherford, K.M.

Nature 409, 1007-1011, 2001

A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; S

A>Title: Massive gene decay in the leprosy bacillus.

A:Reference number: A86909; MUID:21128732; PMID:11234002

A:Accession: B87135

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-320 <STO>

A:Cross-references: GB:AL450380; NID:gl13093520; PIDN:CAC30762.1; GSPDB:GN00147

C:Genetics:

A:Gene: ML1809

Query Match 97.1%; Score 33; DB 2; Length 320;  
 Best Local Similarity 85.7%; Pred. No. 11;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRRGMAI 7  
 DB 197 RRRGMAV 203  
 |||||

RESULT 3  
 T37047  
 probable nitrate reductase delta chain - Streptomyces coelicolor  
 C:Species: Streptomyces coelicolor  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
 C:Accession: T37047  
 R:Murphy, L.; Harris, D.; Thomson, N.R.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
 submitted to the EMBL Data Library, August 1999  
 A:Reference number: Z21619  
 A:Accession: T37047  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-267 <MUR>  
 A:Cross-references: EMBL:AL109889; PIDN:CAB53442.1; GSPDB:GN00070; SCOEDB:SCJ12.30  
 A:Experimental source: strain A3(2)  
 C:Genetics:  
 A:Gene: narX2; SCOEDB:SCJ12.30

Query Match 94.1%; Score 32; DB 2; Length 267;  
 Best Local Similarity 85.7%; Pred. No. 15;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRRGMAI 7  
 DB 127 RRRGMAI 133  
 |||||

RESULT 4  
 W3ADH2  
 late L2 mu core protein precursor - human adenovirus 2  
 C:Species: Mastadenovirus h2 (human adenovirus 2)  
 A:Note: host Homo sapiens (man)  
 C:Date: 30-Sep-1989 #sequence\_revision 30-Jun-1993 #text\_change 31-Dec-1993  
 C:Accession: D03837; A30970  
 R:Alestrom, P.; Akusjaervi, G.; Lager, M.; Yeh-kai, L.; Pettersson, U.  
 J. Biol. Chem. 259, 13980-13985, 1984  
 A:Title: Genes encoding the core proteins of adenovirus type 2.  
 A:Reference number: A03837; MUID:85054835; PMID:6094534  
 A:Accession: D03837  
 A:Molecule type: DNA  
 A:Residues: 1-80 <ALE>  
 A:Cross-references: GB:J01917  
 R:Weber, J.M.; Anderson, C.W.  
 J. Virol. 62, 1741-1745, 1988  
 A:Title: Identification of the gene coding for the precursor of adenovirus core protein  
 A:Reference number: A30970; MUID:88188260; PMID:3357209  
 A:Accession: A30970  
 A:Molecule type: protein  
 A:Residues: 2-80 <WEB>  
 C:Genetics:  
 A:Map position: 49.2-49.9  
 C:Superfamily: adenovirus late L2 mu core protein  
 C:Keywords: core protein; DNA binding; late protein  
 F:32-50/Product: late L2 mu core protein #status experimental <MAT>

Query Match 88.2%; Score 30; DB 1; Length 80;  
 Best Local Similarity 100.0%; Pred. No. 15;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRRGMA 6  
 DB 20 RRRGMA 25  
 |||||

RESULT 5  
 T08768  
 hypothetical protein DKFZp586I111.1 - human (fragment)  
 C:Species: Homo sapiens (man)  
 C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 13-Aug-1999  
 C:Accession: T08768  
 R:Ottewill, B.; Obermaier, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
 submitted to the Protein Sequence Database, May 1999  
 A:Reference number: Z16474  
 A:Accession: T08768  
 A:Molecule type: mRNA  
 A:Residues: 1-185 <OTT>  
 A:Cross-references: EMBL:AL050131  
 A:Experimental source: adult uterus; clone DKFZp586I111  
 C:Genetics:  
 A:Note: DKFZp586I111.1

Query Match 88.2%; Score 30; DB 2; Length 185;  
 Best Local Similarity 100.0%; Pred. No. 31;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRRGMA 6  
 DB 142 RRRGMA 147  
 |||||

RESULT 6  
 A87149  
 conserved hypothetical membrane protein ML1918 [imported] - Mycobacterium leprae  
 C:Species: Mycobacterium leprae  
 C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
 C:Accession: A87149  
 R:Cole, S.T.; Eiglmeyer, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; F.  
 R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd  
 eam, M.A.; Rutherford, K.W.  
 Nature 409, 1007-1011, 2001  
 A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; E  
 A:Title: Massive gene decay in the leprosy bacillus.  
 A:Reference number: A86909; MUID:21128732; PMID:11234002  
 A:Accession: A87149  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-263 <STO>  
 A:Cross-references: GB:AL450380; NID:G13093582; PIDN:CAC30873.1; GSPDB:GN00147  
 C:Genetics:  
 A:Gene: ML1918

Query Match 88.2%; Score 30; DB 2; Length 263;  
 Best Local Similarity 71.4%; Pred. No. 42;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRRGMAI 7  
 DB 17 RRRGLAV 23  
 |||||

RESULT 7  
 H83560  
 probable type II secretion system protein PA0662 [imported] - Pseudomonas aeruginosa (s  
 C:Species: Pseudomonas aeruginosa  
 C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
 C:Accession: H83560  
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; F  
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lin  
 ; Lory, S.; Olson, M.V.  
 Nature 406, 959-964, 2000  
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic path  
 A:Reference number: A82950; MUID:20437337; PMID:10984043  
 A:Accession: H83560  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-321 <STO>  
 A:Cross-references: GB:AE004503; GB:AE004091; NID:G9946553; PIDN:ARG04071.1; GSPDB:GN00



A;Experimental source: strain PA01  
C;Genetics:  
A;Gene: PA0682

Query Match 88.2%; Score 30; DB 2; Length 321;  
Best Local Similarity 85.7%; Pred. No. 50;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRRGMAI 7  
|:|||||  
DB 5 RRRGMAI 11

RESULT 8  
T16625  
asparagine synthase (glutamine-hydrolysing) (EC 6.3.5.4) [similarity] - Caenorhabditis e  
C;Species: Caenorhabditis elegans  
C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 19-Jul-2002  
C;Accession: T16625  
R;Miller, N.  
submitted to the EMBL Data Library, November 1995  
A;Description: The sequence of C. elegans cosmid M02D8.  
A;Reference number: Z18549  
A;Accession: T16625  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-567 <MIL>  
A;Cross-references: EMBL:U41034; NID:g1086742; PID:g1086745; PIDN:AAA82381.1; CBSP:M02D8  
C;Genetics:  
A;Gene: CBSP:M02D8.4  
A;Introns: 28/2; 74/3; 104/3; 161/2; 296/3; 531/1  
C;Superfamily: asparagine synthase (glutamine-hydrolyzing)  
C;Keywords: asparagine biosynthesis; ligase  
F;2/Active site: Cys #status predicted

Query Match 88.2%; Score 30; DB 2; Length 567;  
Best Local Similarity 71.4%; Pred. No. 83;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRRGMAI 7  
|:|||||  
DB 254 KRRGMVA 260

RESULT 9  
H70872  
probable ctpD protein - Mycobacterium tuberculosis (strain H37RV)  
C;Species: Mycobacterium tuberculosis  
C;Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
C;Accession: H70872  
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.  
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A;Reference number: A70500; MUID:98295987; PMID:9634230  
A;Accession: H70872  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-657 <COL>  
A;Cross-references: GB:AL021184; GB:AL123456; NID:g3261498; PIDN:CAA15997.1; PID:g279140  
A;Experimental source: strain H37RV

C;Genetics:  
A;Gene: ctpD  
C;Superfamily: Enterococcus copper-transporting ATPase ctpB; ATPase nucleotide-binding d  
F;25-394/Domain: ATPase transduction domain homology <ATP>  
F;465-608/Domain: ATPase nucleotide-binding domain homology <ATN>

Query Match 88.2%; Score 30; DB 2; Length 657;  
Best Local Similarity 85.7%; Pred. No. 94;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRRGMAI 7  
|:|||||  
DB 399 RRRGMAI 405

# RESULT 10

T37830  
probable glutamate-tRNA ligase (EC 6.1.1.17) - fission yeast (Schizosaccharomyces pombe  
C;Species: Schizosaccharomyces pombe  
C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Jun-2002  
C;Accession: T37830  
R;Devlin, K.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.  
submitted to the EMBL Data Library, August 1997  
A;Reference number: Z21737  
A;Accession: T37830  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-716 <DEV>  
A;Cross-references: EMBL:Z98849; PIDN:CAB11515.1; GSPDB:GN00066; SPDB:SPAC17A5.15c  
A;Experimental source: strain 972h-; cosmid C17A5  
C;Genetics:  
A;Gene: SPDB:SPAC17A5.15c  
A;Map position: 1  
A;Superfamily: yeast glutamate-tRNA ligase; glutamine-tRNA ligase homology  
C;Keywords: ligase

Query Match 88.2%; Score 30; DB 2; Length 716;  
Best Local Similarity 85.7%; Pred. No. 1e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RRRGMAI 7  
|:|||||  
DB 471 RRRGMTI 477

RESULT 11  
T34847  
probable transcription regulator - Streptomyces coelicolor  
C;Species: Streptomyces coelicolor  
C;Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 05-Nov-1999  
C;Accession: T34847  
R;Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, February 1999  
A;Reference number: Z21559  
A;Accession: T34847  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-943 <OLI>  
A;Cross-references: EMBL:AL035478; PIDN:CAB36601.1; GSPDB:GN00070; SCOEDB:SC2G5.14c  
A;Experimental source: strain A3(2)  
C;Genetics:  
A;Gene: SCOEDB:SC2G5.14c

Query Match 88.2%; Score 30; DB 2; Length 943;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRRGMA 6  
|:|||||  
DB 608 RRRGMA 613

# RESULT 12

S07245  
xanthine dehydrogenase (EC 1.1.1.204) - fruit fly (Drosophila melanogaster)  
C;Species: Drosophila melanogaster  
C;Date: 21-Nov-1993 #sequence\_revision 07-Jun-1996 #text\_change 11-Jun-1999  
C;Accession: S07245; S07244; S10132  
R;Keith, T.P.; Riley, M.A.; Kreitman, M.; Lewontin, R.C.; Curtis, D.; Chambers, G.  
Genetics 116, 67-73, 1987  
A;Title: Sequence of the structural gene for xanthine dehydrogenase (rosy locus) in Dro  
A;Reference number: S07245; MUID:87248040; PMID:3036646  
A;Accession: S07245

A:Molecule type: DNA  
A:Residues: 198-1335 <KE1>  
A:Cross-References: EMBL:Y00308  
A>Note: mRNA was also sequenced  
R:Lee, C.S.; Curtis, D.; McCarron, M.; Love, C.; Gray, M.; Bender, W.; Chovnick, A.  
Genetics 116, 55-66, 1987  
A:Title: Mutations affecting expression of the rosy locus in *Drosophila melanogaster*.  
A:Reference number: S07244; MUID:87248039; PMID:3036645  
A:Accession: S07244  
A:Molecule type: DNA  
A:Residues: 1-230 <LEP>  
A:Cross-References: EMBL:Y00308  
A>Note: the authors translated the codon ACC for residue 185 as Ser  
R:Lee, C.S.; Curtis, D.; McCarron, M.; Love, C.; Gray, M.; Bender, W.; Chovnick, A.  
submitted to the EMBL Data Library, February 1987  
A:Reference number: S10132  
A:Accession: S10132  
A:Molecule type: DNA  
A:Residues: 1-105, 'P', 107-1335 <LE2>  
A:Cross-References: EMBL:Y00308; NID:S8830; PIDN:CAA68409.1; PID:G8831  
C:Genetics:  
A:Gene: FlyBase:ry  
A:Cross-References: FlyBase:FBgn000308  
A:Introns: 14/3; 881/3; 1319/3  
C:Superfamily: xanthine dehydrogenase; ferredoxin [2Fe-2S] homology  
C:Keywords: 2Fe-2S; FAD; flavoprotein; iron-sulfur protein; metalloprotein; molybdenum;  
F:26-74/Domain: ferredoxin [2Fe-2S] homology <PER1>  
F:43,48,51,73/Binding site: 2Fe-2S cluster (Cys) (covalent) #status predicted

Query Match 88.2%; Score 30; DB 2; Length 1335;  
Best Local Similarity 71.4%; Pred. No. 1.8e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRRGNMAI 7  
|:|||||  
Db 999 RRRGNMAV 1005

RESULT 13  
A:1946  
xanthine dehydrogenase (EC 1.1.1.204) - fruit fly (*Drosophila pseudoobscura*)  
C:Species: *Drosophila pseudoobscura*  
C:Date: 22-Nov-1989 #sequence\_revision 22-Nov-1989 #text\_change 11-Jun-1999  
C:Accession: A31946  
R:Piley, M.A.  
Mol. Biol. Evol. 6, 33-52, 1989  
A:Title: Nucleotide sequence of the Xdh region in *Drosophila pseudoobscura* and an analysis  
A:Reference number: A31946; MUID:89158785; PMID:2493563  
A:Accession: A31946  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1342 <RIL>  
A:Cross-References: GB:M33977; NID:G158807; PIDN:AAA29022.1; PID:G158809  
C:Genetics:  
A:Gene: FlyBase:Dpse/ry  
A:Cross-References: FlyBase:FBgn0012736  
C:Superfamily: xanthine dehydrogenase; ferredoxin [2Fe-2S] homology  
C:Keywords: 2Fe-2S; FAD; flavoprotein; iron-sulfur protein; metalloprotein; molybdenum;  
F:30-78/Domain: ferredoxin [2Fe-2S] homology <PER1>  
F:47,52,55,77/Binding site: 2Fe-2S cluster (Cys) (covalent) #status predicted

Query Match 88.2%; Score 30; DB 2; Length 1342;  
Best Local Similarity 71.4%; Pred. No. 1.8e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRRGNMAI 7  
|:|||||  
Db 1007 RRRGNMAV 1013

RESULT 14  
A:98305

hypothetical protein AGR\_L2791 [imported] - *Agrobacterium tumefaciens* (strain C58, Cer  
C:Species: *Agrobacterium tumefaciens*  
C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 18-Nov-2002  
A:Accession: A98305  
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman  
A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.  
Science 294, 2323-2328, 2001  
A:Title: Genome sequence of the plant pathogen and biotechnology agent *Agrobacterium tu*  
A:Reference number: A97359; MUID:21608551; PMID:11743194  
A:Accession: A98305  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-62 <KUR>  
A:Cross-References: GB:AE007870; PIDN:AAK89963.1; PID:G15159924; GSPDB:GN00170  
C:Genetics:  
A:Gene: AGR\_L2791  
A:Map position: linear chromosome

Query Match 85.1%; Score 29; DB 2; Length 62;  
Best Local Similarity 85.7%; Pred. No. 19;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RRRGNMAI 7  
|:|||||  
Db 25 RRRGNMAI 31

RESULT 15  
A:2978  
hypothetical protein Atu3429 [imported] - *Agrobacterium tumefaciens* (strain C58, Dupont  
C:Species: *Agrobacterium tumefaciens*  
C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002  
C:Accession: AD2978  
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo,  
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McCiel  
Science 294, 2317-2323, 2001  
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
ster, E.W.  
A:Title: The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58.  
A:Reference number: AB2577; MUID:21608550; PMID:11743193  
A:Accession: AD2978  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-62 <KUR>  
A:Cross-References: GB:AE008689; PIDN:AAL44242.1; PID:G17741825; GSPDB:GN00187  
A:Experimental source: strain C58 (Dupont)  
C:Genetics:  
A:Gene: Atu3429  
A:Map position: linear chromosome

Query Match 85.1%; Score 29; DB 2; Length 62;  
Best Local Similarity 85.7%; Pred. No. 19;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RRRGNMAI 7  
|:|||||  
Db 25 RRRGNMAI 31

Search completed: February 11, 2004, 17:11:36  
Job time : 11.3333 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

QM protein - protein search, using sw model

Run on: February 11, 2004, 17:09:56 ; Search time 24.6667 Seconds  
(without alignments)  
59.419 Million cell updates/sec

Title: US-09-901-187C-4

Perfect score: 34

Sequence: 1 RRRGMAI 7

Scoring table: BLOSUM62  
Gap 10.0, Gapext 0.5

Searched: 801455 seqs, 20938283 residues

Total number of hits satisfying chosen parameters: 801455

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	34	100.0	7	US-09-901-187B-4	Sequence 4, Appli
2	34	100.0	12	US-10-369-493-20777	Sequence 20777, A
3	30	88.2	107	US-10-083-815-71	Sequence 71, Appl
4	30	88.2	112	US-09-864-408A-3560	Sequence 3560, Ap
5	30	88.2	130	US-09-867-550-666	Sequence 666, App
6	30	88.2	263	US-10-080-170-213	Sequence 213, App
7	30	88.2	606	US-10-369-493-13278	Sequence 13278, A
8	30	88.2	618	US-10-369-493-4230	Sequence 4230, Ap
9	30	88.2	715	US-10-128-714-3375	Sequence 3375, Ap
10	30	88.2	715	US-10-128-714-8375	Sequence 8375, Ap
11	30	88.2	716	US-10-369-493-2175	Sequence 2175, Ap
12	29	85.3	129	US-10-131-410-129	Sequence 129, App
13	29	85.3	242	US-10-156-761-13652	Sequence 13652, A
14	29	85.3	264	US-10-080-170-604	Sequence 604, App
15	29	85.3	267	US-10-084-843-71	Sequence 71, Appl

ALIGNMENTS

RESULT 1

US-09-901-187B-4

; Sequence 4, Application US/09901187B

; Patent No. US2002015146A1

; GENERAL INFORMATION:

; APPLICANT: Panacea Pharmaceuticals, Inc.

; APPLICANT: Kolozin, Benjamin

; APPLICANT: Ostretova-Golts, Natalie

; APPLICANT: Lebowitz, Michael S.

; TITLE OF INVENTION: Methods for Preventing Neural Tissue Damage and for the Treatment

; FILE REFERENCE: PAN01/002US

; CURRENT APPLICATION NUMBER: US/09/901,187B

; CURRENT FILING DATE: 2001-07-09

; PRIOR APPLICATION NUMBER: US 60/217,319

; PRIOR FILING DATE: 2000-07-07

; PRIOR APPLICATION NUMBER: US 60/279,199

; PRIOR FILING DATE: 2001-03-28

; NUMBER OF SEQ ID NOS: 12

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 4

; LENGTH: 7

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-901-187B-4

Query Match 100.0%; Score 34; DB 10; Length 7;

Best Local Similarity 100.0%; Pred. No. 7.1e+05;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRRGMAI 7

DB 1 RRRGMAI 7

RESULT 2

US-10-369-493-20777

; Sequence 20777, Application US/10369493

; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 20777  
; LENGTH: 397  
; TYPE: PRT  
; ORGANISM: Rhodospseudomonas palustris  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(397)  
; OTHER INFORMATION: unsure at all Xaa locations  
US-10-369-493-20777

Query Match 100.0%; Score 34; DB 12; Length 397;  
Best Local Similarity 100.0%; Pred. No. 42;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RRRGMAI 7  
Db 126 RRRGMAI 132

RESULT 3  
US-10-083-815-71  
; Sequence 71, Application US/10083815  
; Publication No. US2003026781A1  
; GENERAL INFORMATION:  
; APPLICANT: Anderson, Christen M.  
; APPLICANT: Clevenger, William  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR REGULATING  
; TITLE OF INVENTION: ENDOGENOUS INHIBITOR OF ATP SYNTHASE, INCLUDING  
; TITLE OF INVENTION: TREATMENT FOR DIABETES  
; FILE REFERENCE: 660088.435C2  
; CURRENT APPLICATION NUMBER: US/10/083,815  
; CURRENT FILING DATE: 2002-02-27  
; NUMBER OF SEQ ID NOS: 72  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 71  
; LENGTH: 107  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Fusion protein  
US-10-083-815-71

Query Match 88.2%; Score 30; DB 15; Length 107;  
Best Local Similarity 100.0%; Pred. No. 76;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RRRGMA 6  
Db 45 RRRGMA 50

RESULT 4  
US-09-864-408A-3560  
; Sequence 3560, Application US/09864408A  
; Publication No. US20040009474A1  
; GENERAL INFORMATION:  
; APPLICANT: Leach, Martin D.  
; APPLICANT: Shimkets, Richard A.

; TITLE OF INVENTION: No. US20040009474A1 Human Polynucleotides and Polypeptides Enc  
; FILE REFERENCE: 21402-012  
; CURRENT APPLICATION NUMBER: US/09/864,408A  
; CURRENT FILING DATE: 2001-05-24  
; PRIOR APPLICATION NUMBER: 60/206,690  
; PRIOR FILING DATE: 2000-05-24  
; NUMBER OF SEQ ID NOS: 9068  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3560  
; LENGTH: 112  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-864-408A-3560

Query Match 88.2%; Score 30; DB 12; Length 112;  
Best Local Similarity 71.4%; Pred. No. 79;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RRRGMAI 7  
Db 71 RRRGMAI 77

RESULT 5  
US-09-867-550-666  
; Sequence 666, Application US/09867550  
; Patent No. US20020082206A1  
; GENERAL INFORMATION:  
; APPLICANT: Leach, Martin D.  
; APPLICANT: Mehraban, Fuad,  
; APPLICANT: Conley, Pamela  
; APPLICANT: Law, Debbie  
; APPLICANT: Topper, James  
; TITLE OF INVENTION: No. US20020082206A1 Polynucleotides from Atherogenic Cells and  
; TITLE OF INVENTION: Thereby  
; FILE REFERENCE: 21402-013 (Cura-313)  
; CURRENT APPLICATION NUMBER: US/09/867,550  
; CURRENT FILING DATE: 2001-09-20  
; PRIOR APPLICATION NUMBER: USSN 60/208,427  
; PRIOR FILING DATE: 2000-05-30  
; NUMBER OF SEQ ID NOS: 2125  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 666  
; LENGTH: 130  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-867-550-666

Query Match 88.2%; Score 30; DB 9; Length 130;  
Best Local Similarity 71.4%; Pred. No. 92;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RRRGMAI 7  
Db 20 RRRGMAI 26

RESULT 6  
US-10-080-170-213  
; Sequence 213, Application US/10080170  
; Publication No. US20030129601A1  
; GENERAL INFORMATION:  
; APPLICANT: Cole, S.T.  
; TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR  
; TITLE OF INVENTION: IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR  
; TITLE OF INVENTION: TREATMENT OF MYCOBACTERIOSES  
; FILE REFERENCE: 03495.0218  
; CURRENT APPLICATION NUMBER: US/10/080,170  
; CURRENT FILING DATE: 2002-06-10  
; PRIOR APPLICATION NUMBER: 60/270,123  
; PRIOR FILING DATE: 2001-02-22  
; NUMBER OF SEQ ID NOS: 652  
; SOFTWARE: Patentin Ver. 2.1

; SEQ ID NO 213  
; LENGTH: 263  
; TYPE: PRT  
; ORGANISM: Mycobacterium leprae  
US-10-080-170-213

Query Match 88.2%; Score 30; DB 16; Length 263;  
Best Local Similarity 71.4%; Pred. No. 1.8e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRRGMWAI 7  
Db 17 RRRGLAV 23

RESULT 7  
US-10-369-493-13278  
; Sequence 13278, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIOR FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 13278  
; LENGTH: 606  
; TYPE: PRT  
; ORGANISM: Aspergillus nidulans  
US-10-369-493-13278

Query Match 88.2%; Score 30; DB 12; Length 606;  
Best Local Similarity 85.7%; Pred. No. 4.1e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RRRGMWAI 7  
Db 361 RRRGMTI 367

RESULT 8  
US-10-369-493-4230  
; Sequence 4230, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIOR FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 4230  
; LENGTH: 618  
; TYPE: PRT  
; ORGANISM: Neurospora crassa  
US-10-369-493-4230

Query Match 88.2%; Score 30; DB 12; Length 618;

Best Local Similarity 85.7%; Pred. No. 4.2e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RRRGMWAI 7  
Db 373 RRRGMTI 379

RESULT 9  
US-10-128-714-3375  
; Sequence 3375, Application US/10128714  
; Publication No. US20030119013A1  
; GENERAL INFORMATION:  
; APPLICANT: Jiang, Bo  
; APPLICANT: Hu, Wengqi  
; APPLICANT: Tishkoff, Daniel  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Ershkin, Alexey M  
; APPLICANT: Lemieux, Sebastien M  
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and  
; FILE REFERENCE: 10182-018-999  
; CURRENT APPLICATION NUMBER: US/10/128,714  
; CURRENT FILING DATE: 2002-04-23  
; PRIOR APPLICATION NUMBER: US 60/285,697  
; PRIOR FILING DATE: 2001-04-23  
; PRIOR APPLICATION NUMBER: US 60/287,066  
; PRIOR FILING DATE: 2001-04-27  
; PRIOR APPLICATION NUMBER: US 60/295,890  
; PRIOR FILING DATE: 2001-06-05  
; PRIOR APPLICATION NUMBER: US 60/303,899  
; PRIOR FILING DATE: 2001-07-09  
; PRIOR APPLICATION NUMBER: US 60/316,362  
; PRIOR FILING DATE: 2001-08-31  
; NUMBER OF SEQ ID NOS: 8603  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 3375  
; LENGTH: 715  
; TYPE: PRT  
; ORGANISM: Aspergillus fumigatus  
US-10-128-714-3375

Query Match 88.2%; Score 30; DB 15; Length 715;  
Best Local Similarity 85.7%; Pred. No. 4.8e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RRRGMWAI 7  
Db 473 RRRGMTI 479

RESULT 10  
US-10-128-714-8375  
; Sequence 8375, Application US/10128714  
; Publication No. US20030119013A1  
; GENERAL INFORMATION:  
; APPLICANT: Jiang, Bo  
; APPLICANT: Hu, Wengqi  
; APPLICANT: Tishkoff, Daniel  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Ershkin, Alexey M  
; APPLICANT: Lemieux, Sebastien M  
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and  
; FILE REFERENCE: 10182-018-999  
; CURRENT APPLICATION NUMBER: US/10/128,714  
; CURRENT FILING DATE: 2002-04-23  
; PRIOR APPLICATION NUMBER: US 60/285,697  
; PRIOR FILING DATE: 2001-04-23  
; PRIOR APPLICATION NUMBER: US 60/287,066  
; PRIOR FILING DATE: 2001-04-27  
; PRIOR APPLICATION NUMBER: US 60/295,890  
; PRIOR FILING DATE: 2001-06-05

;; PRIOR APPLICATION NUMBER: US 60/303,899  
;; PRIOR FILING DATE: 2001-07-09  
;; PRIOR APPLICATION NUMBER: US 60/316,362  
;; PRIOR FILING DATE: 2001-08-31  
;; NUMBER OF SEQ ID NOS: 8603  
;; SOFTWARE: PatentIn version 3.1  
;; SEQ ID NO 8375  
;; LENGTH: 715  
;; TYPE: PRT  
;; ORGANISM: Aspergillus fumigatus  
US-10-128-714-8375

Query Match 88.2%; Score 30; DB 15; Length 715;  
Best Local Similarity 85.7%; Pred. No. 4.8e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RRRGMAT 7  
DB 473 RRRGMTI 479

## RESULT 11

US-10-369-493-2175  
;; Sequence 2175, Application US/10369493  
;; Publication No. US20030233675A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Cao, Yongwei  
;; APPLICANT: Hinkle, Gregory J.  
;; APPLICANT: Slater, Steven C.  
;; APPLICANT: Goldman, Barry S.  
;; APPLICANT: Chen, Xianfeng  
;; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
;; FILE REFERENCE: 38-10(52052)/B  
;; CURRENT APPLICATION NUMBER: US/10/369,493  
;; CURRENT FILING DATE: 2003-02-28  
;; PRIOR APPLICATION NUMBER: US 60/360,039  
;; PRIOR FILING DATE: 2002-02-21  
;; NUMBER OF SEQ ID NOS: 47374  
;; SEQ ID NO 2175  
;; LENGTH: 716  
;; TYPE: PRT  
;; ORGANISM: Schizosaccharomyces pombe  
US-10-369-493-2175

Query Match 88.2%; Score 30; DB 12; Length 716;  
Best Local Similarity 85.7%; Pred. No. 4.8e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RRRGMAT 7  
DB 471 RRRGMTI 477

## RESULT 12

US-10-131-410-129  
;; Sequence 129, Application US/10131410  
;; Publication No. US20030235915A1  
;; GENERAL INFORMATION:  
;; APPLICANT: SPECHT, THOMAS  
;; APPLICANT: HINZMANN, BERND  
;; APPLICANT: SCHMITT, ARMIN  
;; APPLICANT: PILARSKY, CHRISTIAN  
;; APPLICANT: DAHL, EDGAR  
;; APPLICANT: ROSENTHAL, ANDRE  
;; TITLE OF INVENTION: HUMAN NUCLEIC ACID SEQUENCES FROM TISSUE OF BREAST  
;; FILE REFERENCE: SCH-1763  
;; CURRENT APPLICATION NUMBER: US/10/131,410  
;; CURRENT FILING DATE: 2002-04-25  
;; PRIOR APPLICATION NUMBER: 09/646,673  
;; PRIOR FILING DATE: 2000-09-20  
;; PRIOR APPLICATION NUMBER: PCT/DE99/00908

;; PRIOR FILING DATE: 1999-03-19  
;; NUMBER OF SEQ ID NOS: 202  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO 129  
;; LENGTH: 129  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-10-131-410-129

Query Match 85.3%; Score 29; DB 12; Length 129;  
Best Local Similarity 71.4%; Pred. No. 1.5e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRRGMAT 7  
DB 98 RRRGMAT 104

## RESULT 13

US-10-156-761-13652  
;; Sequence 13652, Application US/10156761  
;; Publication No. US20030119018A1  
;; GENERAL INFORMATION:  
;; APPLICANT: OMURA, SATOSHI  
;; APPLICANT: IKEDA, HARUO  
;; APPLICANT: ISHIKAWA, JUN  
;; APPLICANT: HORIKAWA, HIROSHI  
;; APPLICANT: SHIBA, TADAYOSHI  
;; APPLICANT: SAKAKI, YOSHIYUKI  
;; APPLICANT: HAKTORI, MASAHIRA  
;; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
;; FILE REFERENCE: 249-262  
;; CURRENT APPLICATION NUMBER: US/10/156,761  
;; CURRENT FILING DATE: 2002-05-29  
;; PRIOR APPLICATION NUMBER: JP 2001-204089  
;; PRIOR FILING DATE: 2001-05-30  
;; PRIOR APPLICATION NUMBER: JP 2001-272697  
;; PRIOR FILING DATE: 2001-08-02  
;; NUMBER OF SEQ ID NOS: 15109  
;; SEQ ID NO 13652  
;; LENGTH: 242  
;; TYPE: PRT  
;; ORGANISM: Streptomyces avermitilis  
US-10-156-761-13652

Query Match 85.3%; Score 29; DB 15; Length 242;  
Best Local Similarity 71.4%; Pred. No. 2.7e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRRGMAT 7  
DB 93 RRRGLAL 99

## RESULT 14

US-10-080-170-604  
;; Sequence 604, Application US/10080170  
;; Publication No. US20030129601A1  
;; GENERAL INFORMATION:  
;; APPLICANT: COLE, S.T.  
;; TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR  
;; TITLE OF INVENTION: IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR  
;; TITLE OF INVENTION: TREATMENT OF MYCOBACTERIOSES  
;; FILE REFERENCE: 03495.0218  
;; CURRENT APPLICATION NUMBER: US/10/080,170  
;; CURRENT FILING DATE: 2002-06-10  
;; PRIOR APPLICATION NUMBER: 60/270,123  
;; PRIOR FILING DATE: 2001-02-22  
;; NUMBER OF SEQ ID NOS: 652  
;; SOFTWARE: PatentIn Ver. 2.1  
;; SEQ ID NO 604  
;; LENGTH: 264  
;; TYPE: PRT

FT TRANSMEM 86 106 POTENTIAL.  
 FT DOMAIN 107 258 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).  
 FT METAL 186 186 COPPER A (PROBABLE).  
 FT METAL 221 221 COPPER A (PROBABLE).  
 FT METAL 225 225 COPPER A (PROBABLE).  
 FT METAL 229 229 COPPER A (PROBABLE).  
 SQ SEQUENCE 258 AA; 29001 MW; FD514B00A9524A59 CRC64;

Query Match 87.8%; Score 36; DB 1; Length 258;  
 Best Local Similarity 85.7%; Pred. No. 4.4;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HYAKNPI 7  
 DB 64 HYKKNPI 70

RESULT 2  
 COX2\_BETVU STANDARD; PRT; 260 AA.

AC P98012;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Cytochrome c oxidase polypeptide II (EC 1.9.3.1).  
 GN COX2 OR COXII.  
 OS Beta vulgaris (Sugar beet).  
 OG Mitochondrion.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Caryophyllales; Caryophyllales; Chenopodiaceae; Beta.  
 OC NCBI\_TaxID=161934;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV\_NB-1; TISSUE=Root;  
 RX MEDLINE=9135954; PubMed=1653062;  
 RA Mann V., Ekstein I., Nissen H., Hiser C., McIntosh L., Hirschberg J.;  
 RT "The cytochrome oxidase II gene in mitochondria of the sugar-beet  
 Beta vulgaris L."; Plant Mol. Biol. 17:559-566(1991).  
 RL  
 CC -!- FUNCTION: Cytochrome c oxidase is the component of the respiratory  
 chain that catalyzes the reduction of oxygen to water. Subunits 1-  
 3 form the functional core of the enzyme complex. Subunit 2  
 transfers the electrons from cytochrome c via its binuclear copper  
 A center to the bimetallic center of the catalytic subunit 1.  
 CC -!- CATALYTIC ACTIVITY: 4 ferrocycytochrome c + O(2) = 4 ferricytochrome  
 c + 2 H(2)O.  
 CC -!- COFACTOR: Copper A.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial  
 inner membrane.  
 CC -!- SIMILARITY: Belongs to the cytochrome c oxidase subunit 2 family.  
 CC -!- CAUTION: Due to conserved RNA editing events, the following  
 positions might change: 13, 54, 85, 127, 148, 154, 159, 194 and  
 248.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 EMBL; X55297; CAA39009.1;  
 DR InterPro; IPR001505; Copper\_CuA.  
 DR InterPro; IPR002429; Cyt\_c\_ox\_2.  
 DR Pfam; PF00116; COX2; 1.  
 DR Pfam; PF02790; COX2\_TM; 1.  
 DR PRINTS; PR01166; CYCOXIDASEII.  
 DR ProDom; PD000131; Copper\_CuA; 1.  
 DR PROSITE; PS00078; COX2; 1.  
 KW Oxidoreductase; Copper; Mitochondrion; Transmembrane; Inner membrane;  
 Electron transport; Respiratory chain; RNA editing.

FT DOMAIN 1 41  
 FT TRANSMEM 42 58  
 FT DOMAIN 59 82  
 FT TRANSMEM 83 104  
 FT DOMAIN 105 260  
 FT METAL 187 187  
 FT METAL 222 222  
 FT METAL 226 226  
 FT METAL 230 230  
 SQ SEQUENCE 260 AA; 29520 MW; 0AA70D908A866D8 CRC64;

Query Match 87.8%; Score 36; DB 1; Length 260;  
 Best Local Similarity 85.7%; Pred. No. 4.4;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HYAKNPI 7  
 DB 65 HYKKNPI 71

RESULT 3  
 COX2\_SOYBN STANDARD; PRT; 260 AA.

AC P05451;  
 DT 01-NOV-1988 (Rel. 09, Created)  
 DT 01-NOV-1988 (Rel. 09, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Cytochrome c oxidase polypeptide II (EC 1.9.3.1).  
 GN COX2 OR COXII OR COII.  
 OS Glycine max (soybean).  
 OG Mitochondrion.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eucosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.  
 OC NCBI\_TaxID=3847;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88194682; PubMed=2834094;  
 RA Grabau E.A.;  
 RT "Cytochrome oxidase subunit II gene is adjacent to an initiator  
 methionine tRNA gene in soybean mitochondrial DNA."; Curr. Genet. 11:287-293(1987).  
 RL  
 CC -!- FUNCTION: Cytochrome c oxidase is the component of the respiratory  
 chain that catalyzes the reduction of oxygen to water. Subunits 1-  
 3 form the functional core of the enzyme complex. Subunit 2  
 transfers the electrons from cytochrome c via its binuclear copper  
 A center to the bimetallic center of the catalytic subunit 1.  
 CC -!- CATALYTIC ACTIVITY: 4 ferrocycytochrome c + O(2) = 4 ferricytochrome  
 c + 2 H(2)O.  
 CC -!- COFACTOR: Copper A.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial  
 inner membrane.  
 CC -!- SIMILARITY: THE GENE CODING FOR THIS PROTEIN IS PROBABLY  
 SILENT. SOYBEAN NUCLEAR GENOME ENCODE AN ACTIVE COX2.  
 CC -!- SIMILARITY: Belongs to the cytochrome c oxidase subunit 2 family.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 EMBL; X04825; CAA28508.1;  
 DR InterPro; IPR001505; Copper\_CuA.  
 DR InterPro; IPR002429; Cyt\_c\_ox\_2.  
 DR Pfam; PF00116; COX2; 1.  
 DR Pfam; PF02790; COX2\_TM; 1.  
 DR PRINTS; PR01166; CYCOXIDASEII.  
 DR ProDom; PD000131; Copper\_CuA; 1.  
 DR PROSITE; PS00078; COX2; 1.  
 KW Oxidoreductase; Copper; Mitochondrion; Transmembrane; Inner membrane;  
 Electron transport; Respiratory chain; RNA editing.

Query Match 82.9%; Score 34; DB 1; Length 551;  
Best Local Similarity 93.3%; Pred. No. 23;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HYAKNP 6  
DB 145 HYSKNP 150  
|||  
|||

RESULT 5  
COX2 MARPO STANDARD; PRT; 251 AA.  
ID COX2 MARPO  
AC P26857;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Cytochrome c oxidase polypeptide II (EC 1.9.3.1).  
GN COX2 OR COXII.  
OS Marchantia polymorpha (Liverwort).  
OC Mitochondrion.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Marchantiophyta; Marchantiaceae;  
OC Marchantiopsida; Marchantiidae; Marchantiales; Marchantiaceae;  
OC Marchantiaceae; Marchantia.  
OX NCBI\_TaxID=3197;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92114051; PubMed=1731062;  
RA Oda K., Yamato K., Ohta E., Nakamura Y., Takemura M., Nozato N.,  
RA Akashi K., Kanegae T., Ogura Y., Kohchi T., Ohyama K.;  
RT "Gene organization deduced from the complete sequence of liverwort  
Marchantia polymorpha mitochondrial DNA. A primitive form of plant  
mitochondrial genome.";  
RL J. Mol. Biol. 223:1-7(1992).  
CC -!- FUNCTION: Cytochrome c oxidase is the component of the respiratory  
chain that catalyzes the reduction of oxygen to water. Subunits 1-  
3 form the functional core of the enzyme complex. Subunit 2  
transfers the electrons from cytochrome c via its binuclear copper  
A center to the binuclear center of the catalytic subunit 1.  
CC -!- CATALYTIC ACTIVITY: 4 ferrocyanide + O(2) = 4 ferricyanide  
CC c + 2 H(2)O.  
CC -!- COFACTOR: Copper A.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial  
inner membrane.  
CC -!- SIMILARITY: Belongs to the cytochrome c oxidase subunit 2 family.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; M68929; AAC09432.1; -.  
CC InterPro; IPR001505; Copper\_Cua.  
DR InterPro; IPR002429; Cyt\_c\_ox\_2.  
DR Pfam; PF00116; COX2; 1.  
DR Pfam; PF02790; COX2 TM; 1.  
DR PRINTS; PR01166; CYCOXIDASEII.  
DR ProDom; PD000131; Copper Cua; 1.  
DR PROSITE; PS00078; COX2; 1.  
KW Oxidoreductase; Copper; Mitochondrion; Transmembrane; Inner membrane;  
KW Electron transport; Respiratory chain.  
FT DOMAIN 1 39 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).  
FT TRANSMEM 1 39 POTENTIAL.  
FT TRANSMEM 40 60 POTENTIAL.  
FT TRANSMEM 61 75 POTENTIAL.  
FT TRANSMEM 76 104 POTENTIAL.  
FT TRANSMEM 105 251 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL).  
FT METAL 184 184 COPPER A (PROBABLE).  
FT METAL 219 219 COPPER A (PROBABLE).  
FT METAL 223 223 COPPER A (PROBABLE).  
FT METAL 227 227 COPPER A (PROBABLE).  
SQ SEQUENCE 251 AA; 28893 MW; 1A6527DB9503278A CRC64;

Query Match 87.8%; Score 36; DB 1; Length 260;  
Best Local Similarity 85.7%; Pred. No. 4.4;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HYAKNPI 7  
DB 64 HYKKNPI 70  
|||  
|||

RESULT 4  
Y275 HAEIN STANDARD; PRT; 551 AA.  
ID Y275 HAEIN  
AC P43975;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein H10275.  
GN H10275.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
OC Pasteurellaceae; Haemophilus.  
OX NCBI\_TaxID=727;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=RD / KW20 / ATCC 51907;  
RX MEDLINE=95350630; PubMed=7542800;  
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,  
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
RA Venter J.C.;  
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
Rd.";  
RL Science 269:496-512(1995).  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; U32714; AAC21949.1; -.  
DR PIR; B64005; B64005.  
DR TIGR; H10275; -.  
KW Hypothetical protein; Transmembrane; Complete proteome.  
FT TRANSMEM 1 21 POTENTIAL.  
FT TRANSMEM 25 45 POTENTIAL.  
FT TRANSMEM 99 119 POTENTIAL.  
FT TRANSMEM 124 144 POTENTIAL.  
FT TRANSMEM 266 286 POTENTIAL.  
FT TRANSMEM 490 510 POTENTIAL.  
SQ SEQUENCE 551 AA; 62256 MW; BF08FE215695D610 CRC64;



Query Match 80.5%; Score 33; DB 1; Length 251;  
 Best Local Similarity 71.4%; Pred. No. 17;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 HYAKNPI 7  
 |||||  
 Db 63 HYKKNPM 69

RESULT 6  
 PTNB MOUSE  
 ID PTNB MOUSE STANDARD; PRT; 585 AA.  
 AC P35235;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Protein-tyrosine phosphatase, non-receptor type 11 (EC 3.1.3.48)  
 DE (Protein-tyrosine phosphatase SYP).  
 GN PTPN11.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=93206094; PubMed=8096088;  
 RX Feng G.-S., Hui C.-C., Pawson T.;  
 RA "SH2-containing phosphotyrosine phosphatase as a target of protein-  
 tyrosine kinases.";  
 RL Science 259:1607-1611 (1993).  
 RN [2]  
 RN PTPNS1 BINDING.  
 RP MEDLINE=97215901; PubMed=9062191;  
 RX Kharitonov A., Chen Z., Sures I., Wang H., Schilling J.,  
 RA Ullrich A.;  
 RA "A family of proteins that inhibit signalling through tyrosine kinase  
 receptors.";  
 RL Nature 386:181-186 (1997).  
 RN [3]  
 RN X-RAY CRYSTALLOGRAPHY (2.05 ANGSTROMS) OF 4-103.  
 RX MEDLINE=94363243; PubMed=7521735;  
 RA Lee C.-H., Kominos D., Jacques S., Margolis B., Schlessinger J.,  
 RA Shoenelson S.E., Kuriyan J.;  
 RA "Crystal structures of peptide complexes of the amino-terminal SH2  
 domain of the SYP tyrosine phosphatase.";  
 RL Structure 2:423-438 (1994).  
 CC -!- FUNCTION: This PTPase activity may directly link growth factor  
 receptors and other signaling proteins through protein-tyrosine  
 phosphorylation. The SH2 regions may interact with other cellular  
 components to modulate its own phosphatase activity against  
 interacting substrates.  
 CC -!- CATALYTIC ACTIVITY: Protein tyrosine phosphatase + H(2)O = protein  
 tyrosine + phosphate.  
 CC -!- SUBUNIT: Binds PTPNS1.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -!- PTM: Phosphorylated by tyrosine-protein kinases.  
 CC -!- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-  
 TYROSINE PHOSPHATASE FAMILY. SURCLASS THAT CONTAINS SH2 DOMAINS.  
 CC -!- SIMILARITY: Contains 2 SH2 domains.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL curators on  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; L08663; ; NOT ANNOTATED\_CDS.  
 PIR; A46209; A46209.  
 PDB; 1AYC; 31-AUG-94.  
 PDB; 1AYB; 31-AUG-94.

DR PDB; 1AYC; 31-AUG-94.  
 DR PDB; 1AYD; 31-AUG-94.  
 DR MGD; MGI:99511; Ptpn11.  
 DR GO; CO:0007409; P-axogenesis; IMP.  
 DR GO; GO:0048011; P-NGF receptor signaling pathway; IMP.  
 DR InterPro; IPR000980; SH2.  
 DR InterPro; IPR000387; TYR phosphatase.  
 DR InterPro; IPR000242; Tyr\_PP.  
 DR Pfam; PF00102; Y-phosphatase; 1.  
 DR PRINTS; PR00700; PRTYVPHPTASE.  
 DR PRINTS; PR00401; SH2DOMAIN.  
 DR PRODOM; PD000093; SH2; 2.  
 DR SMART; SM00194; PTPC; 1.  
 DR SMART; SM00252; SH2; 2.  
 DR PROSITE; PS00383; TYR\_PHOSPHATASE\_1; 1.  
 DR PROSITE; PS00056; TYR\_PHOSPHATASE\_2; 1.  
 DR PROSITE; PS00055; TYR\_PHOSPHATASE\_PTP; 1.  
 DR PROSITE; PS00001; SH2; 2.  
 DR Hydrolase; SH2 domain; Repeat; Phosphorylation; 3D-structure.  
 KW SH2 1.  
 FT DOMAIN 6 102  
 FT DOMAIN 112 216  
 FT DOMAIN 276 521  
 FT ACT\_SITE 463 463  
 FT TURN 5 6  
 FT TURN 7 7  
 FT STRAND 13 22  
 FT HELIX 23 23  
 FT TURN 26 27  
 FT TURN 28 33  
 FT STRAND 38 39  
 FT TURN 41 47  
 FT STRAND 48 49  
 FT TURN 50 55  
 FT STRAND 57 58  
 FT STRAND 63 64  
 FT STRAND 71 71  
 FT STRAND 74 83  
 FT HELIX 85 86  
 FT TURN 89 90  
 FT STRAND 91 92  
 FT TURN 95 95  
 FT STRAND 100 101  
 FT STRAND 100 101  
 SQ SEQUENCE 585 AA; 66816 MW; 6CE554F929B8F72A CRC64;

Query Match 80.5%; Score 33; DB 1; Length 585;  
 Best Local Similarity 71.4%; Pred. No. 39;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 HYAKNPI 7  
 |||||  
 Db 196 HYKKNPM 202

RESULT 7  
 PTNB HUMAN  
 ID PTNB HUMAN STANDARD; PRT; 593 AA.  
 AC Q06124;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Protein-tyrosine phosphatase, non-receptor type 11 (EC 3.1.3.48)  
 DE (Protein-tyrosine phosphatase 2C) (PTP-2C) (SH-PTP3) (SH-  
 PTP2) (SHP-2).  
 GN PTPN11 OR PTP2C OR SHPTP2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RP TISSUE=Umbilical cord;  
 RC MEDLINE=93211929; PubMed=7681589;  
 RX

RA Ahmad S., Banville D.L., Zhao Z., Fischer E.H., Shen S.H.;  
 RT "A widely expressed human protein-tyrosine phosphatase containing src  
 RT homology 2 domains.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:2197-2201(1993).  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=93206095; PubMed=7661217;  
 RX Vogel W., Lammers R., Huang J., Ullrich A.;  
 RA "Activation of a phosphotyrosine phosphatase by tyrosine  
 RT phosphorylation.";  
 RL Science 259:1611-1614(1993).  
 RN [3]  
 RN SEQUENCE FROM N.A.  
 RP TISSUE=T-cell;  
 RX MEDLINE=93106179; PubMed=1281790;  
 RA Adachi M., Sekiya M., Miyachi T., Mateuno K., Hinoda Y., Imai K.,  
 RA Tachi A.;  
 RT "Molecular cloning of a novel protein-tyrosine phosphatase SH-PTP3  
 RT with sequence similarity to the src-homology region 2.";  
 RL FEBS Lett. 314:335-339(1992).  
 RN [4]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=94029983; PubMed=8216283;  
 RX Bastien L., Ramachandran C., Liu S., Adam M.;  
 RA "Cloning, expression and mutational analysis of SH-PTP2, human  
 RT protein-tyrosine phosphatase.";  
 RL Biochem. Biophys. Res. Commun. 196:124-133(1993).  
 RN [5]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=93087502; PubMed=1280823;  
 RX Freeman R.M., Jr., Plutsky J., Neel B.G.;  
 RA "Identification of a human src homology 2-containing protein-tyrosine-  
 RT phosphatase: a putative homolog of Drosophila corkscrew.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:11239-11243(1992).  
 RN [6]  
 RN PHOSPHORYLATION BY PDGFR.  
 RP MEDLINE=94316690; PubMed=8041791;  
 RX Bennett A.M., Tang T.L., Sugimoto S., Walsh C.T., Neel B.G.;  
 RA "Protein-tyrosine-phosphatase SHPTP2 couples platelet-derived growth  
 RT factor receptor beta to Ras.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:7335-7339(1994).  
 RN [7]  
 RN INTERACTION WITH PTPNS1.  
 RP MEDLINE=97215901; PubMed=9062191;  
 RX Khaitonenkov A., Chen Z., Sures I., Wang H., Schilling J.,  
 RA Ullrich A.;  
 RT "A family of proteins that inhibit signalling through tyrosine kinase  
 RT receptors.";  
 RL Nature 386:181-186(1997).  
 RN [8]  
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 1-526.  
 RX MEDLINE=98150850; PubMed=9491886;  
 RA Hof P., Pluskey S., Dne-Paganon S., Eck M.J., Shoelson S.E.;  
 RT "Crystal structure of the tyrosine phosphatase SHP-2.";  
 RL Cell 92:441-450(1998).  
 RN [9]  
 RN VARIANTS NS GLY-61; CYS-63; GLY-72; SER-72; ASP-76; ARG-79; VAL-282;  
 RP ASP-308 AND VAL-504.  
 RX MEDLINE=21583743; PubMed=11704759;  
 RA Tartaglia M., Mehler E.L., Goldberg R., Zampino G., Brunner H.G.,  
 RA Kremer H., van der Burg I., Crosby A.H., Ion A., Jeffery S.,  
 RA Kalidas K., Patton M.A., Kucherlapati R.S., Gelb B.D.;  
 RT "Mutations in PTPN11, encoding the protein tyrosine phosphatase SHP-2,  
 RT cause Noonan syndrome.";  
 RL Nat. Genet. 29:465-468(2001).  
 RN [10]  
 RN ERRATUM.  
 RA Tartaglia M., Mehler E.L., Goldberg R., Zampino G., Brunner H.G.,  
 RA Kremer H., van der Burg I., Crosby A.H., Ion A., Jeffery S.,  
 RA Kalidas K., Patton M.A., Kucherlapati R.S., Gelb B.D.;  
 RL Nat. Genet. 29:491-491(2001).  
 RN [11]  
 RN ERRATUM.

RA Tartaglia M., Mehler E.L., Goldberg R., Zampino G., Brunner H.G.,  
 RA Kremer H., van der Burg I., Crosby A.H., Ion A., Jeffery S.,  
 RA Kalidas K., Patton M.A., Kucherlapati R.S., Gelb B.D.;  
 RL Nat. Genet. 30:123-123(2001).  
 RN [12]  
 RN VARIANTS NS ALA-42; ALA-60; ASN-61; GLY-61; ASP-62; CYS-63; GLY-72;  
 RP ILE-73; ASP-76; ARG-79; ALA-106; ASP-139; CYS-279; VAL-282; LEU-285;  
 RX SER-285; ASP-308; SER-308; VAL-309; LYS-501 AND VAL-504, AND VARIANT  
 RP NOONAN-LIKE SYNDROME SER-308.  
 RX MEDLINE=21987645; PubMed=11992261;  
 RA Tartaglia M., Kalidas K., Shaw A., Song X., Musat D.L.,  
 RA van der Burg I., Brunner H.G., Bertola D.R., Crosby A., Ion A.,  
 RA Kucherlapati R.S., Jeffery S., Patton M.A., Gelb B.D.;  
 RT "PTPN11 mutations in Noonan syndrome: molecular spectrum, genotype-  
 RT phenotype correlation, and phenotypic heterogeneity.";  
 RL Am. J. Hum. Genet. 70:1555-1563(2002).  
 RN [13]  
 RN VARIANTS LEOPARD SYNDROME CYS-279 AND MET-468.  
 RP MEDLINE=22104852; PubMed=12058348;  
 RX Digilio M.C., Conti E., Sarkozy A., Mingarelli R., Dottorini T.,  
 RA Maximo B., Pizzuti A., Dallapiccola B.;  
 RT "Grouping of multiple-lentiginos/LEOPARD and Noonan syndromes on the  
 RT PTPN11 gene.";  
 RL Am. J. Hum. Genet. 71:389-394(2002).  
 RN [14]  
 RN VARIANTS NS ASP-62; CYS-63 AND THR-502.  
 RP MEDLINE=22236043; PubMed=12325025;  
 RX Maheshwari M., Belmont J., Fernbach S., Ho T., Molinari L., Yakub I.,  
 RA Yu F., Combes A., Towbin J., Craigen W.J., Gibbs R.;  
 RT "PTPN11 mutations in Noonan syndrome type I: detection of recurrent  
 RT mutations in exons 3 and 13.";  
 RL Hum. Mutat. 20:298-304(2002).  
 RN [15]  
 RN VARIANTS NS GLY-61; CYS-63; SER-72; ILE-73; SER-285 AND ASP-308.  
 RP MEDLINE=22151235; PubMed=12161469;  
 RX Kosaki K., Suzuki T., Muroya K., Hasegawa T., Sato S., Matsuo N.,  
 RA Kosaki R., Nagai T., Hasegawa Y., Ogata T.;  
 RT "PTPN11 (protein-tyrosine phosphatase, nonreceptor-type 11) mutations  
 RT in seven Japanese patients with Noonan syndrome.";  
 RL J. Clin. Endocrinol. Metab. 87:3529-3533(2002).  
 CC -1- FUNCTION: This PTPase activity may directly link growth factor  
 CC receptors and other signaling proteins through protein-tyrosine  
 CC phosphorylation. The SH2 regions may interact with other cellular  
 CC components to modulate its own phosphatase activity against  
 CC interacting substrates.  
 CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O = protein  
 CC tyrosine + phosphate.  
 CC -1- SUBUNIT: Binds PTPNS1.  
 CC -1- SUBCELLULAR LOCATION: Widely expressed; particularly abundant in  
 CC heart, brain, and skeletal muscle.  
 CC -1- PTM: Phosphorylation of tyrosine residues at the C-terminus by  
 CC platelet-derived growth factor creates a binding site for the SH2  
 CC domain of GRB2.  
 CC -1- DISEASE: Defects in PTPN11 are the cause of LEOPARD syndrome  
 CC [MIM:151100], an autosomal dominant disorder allelic with Noonan  
 CC syndrome. The acronym LEOPARD stands for lentiginos,  
 CC electrocardiographic conduction abnormalities, ocular  
 CC hypertelorism, pulmonic stenosis, abnormalities of genitalia,  
 CC retardation of growth, and deafness.  
 CC -1- DISEASE: Defects in PTPN11 are a cause of Noonan syndrome (NS)  
 CC [MIM:163950]; also designated Noonan syndrome 1 (NS1). NS is an  
 CC autosomal dominant disorder characterized by dysmorphic facial  
 CC features, short stature, hypertelorism, cardiac anomalies,  
 CC deafness, motor delay, and a bleeding diathesis. It is a  
 CC genetically heterogeneous and relatively common syndrome, with an  
 CC estimated incidence of 1 in 1000-2500 live births. Mutations in  
 CC PTPN11 account for more than 50% of the cases.  
 CC -1- DISEASE: Defects in PTPN11 are a cause of Noonan-like syndrome  
 CC [MIM:163955]; also known as Noonan-like/multiple giant cell lesion  
 CC syndrome. It is an autosomal dominant disorder characterized by  
 CC Noonan features associates with giant cell lesions of bone and  
 CC soft tissue.

```

CC -!- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
CC TYROSINE PHOSPHATASE FAMILY. SUBCLASS THAT CONTAINS SH2 DOMAINS.
CC
CC -!- SIMILARITY: Contains 2 SH2 domains.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; L08807; -; NOT ANNOTATED_CDS.
CC EMBL; X70766; CAA50045.1; -
CC EMBL; D13540; BAA02740.2; -
CC EMBL; L07527; AAA17022.1; -
CC EMBL; L03535; AAA36611.1; -
CC PIR; JN0805; JN0805.
CC PDB; 2SHF; 16-FEB-99.
CC Genew; HGNC:9644; PTPN11.
CC MIM; 176876; -
CC MIM; 151100; -
CC MIM; 163950; -
CC MIM; 163955; -
CC GO; GO:0004726; F:non-membrane spanning protein tyrosine phos. . . ; TAS.
CC InterPro; IPR000980; SH2.
CC InterPro; IPR000387; TYR_phosphatase.
CC InterPro; IPR000242; Tyr_PP.
CC Pfam; PF00017; SH2; 2.
CC Pfam; PF00102; Y_phosphatase; 1.
CC PRINTS; PR00700; PRTYPHPTASE.
CC PRINTS; PR00401; SH2DOMAIN.
CC PRODOM; PD000093; SH2; 2.
CC SMART; SM00194; PTPC; 1.
CC SMART; SM00252; SH2; 2.
CC PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
CC PROSITE; PS00383; TYR_PHOSPHATASE_2; 1.
CC PROSITE; PS00056; TYR_PHOSPHATASE_2; 1.
CC
CC Query Match 80.5%; Score 33; DB 1; Length 593;
CC Best Local Similarity 71.4%; Pred. No. 40;
CC Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
CC
CC QY 1 HYAKNPI 7
CC |||||
CC Db 196 HYKNPM 202
CC
CC RESULT 8
CC FTN8 RAT STANDARD; PRT; 593 AA.
CC AC P41499; Q62626;
CC DT 01-NOV-1995 (Rel. 32, Created)
CC DT 15-JUL-1998 (Rel. 36, Last sequence update)
CC DT 15-SEP-2003 (Rel. 42, Last annotation update)
CC DE Protein-tyrosine phosphatase, non-receptor type 11 (EC 3.1.3.48)
CC DE (Protein-tyrosine phosphatase SYP).
CC GN PTPN11.
CC OS Rattus norvegicus (Rat).
CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
CC OC NCBI_TaxID=10116;
CC RN [1]
CC SEQUENCE FROM N.A.
CC RC STRAIN-Sprague-Dawley;
CC RX MEDLINE=94324984; PubMed=8048963;
CC RA Ding W., Zhang W.R., Sullivan K., Hashimoto N., Goldstein B.J.;
CC RT "Identification of protein-tyrosine phosphatases prevalent in
CC RT adipocytes by molecular cloning."
CC RL Biochem. Biophys. Res. Commun. 202:902-907(1994).
CC RN [2]
CC SEQUENCE FROM N.A.
CC RP STRAIN-Sprague-Dawley;
CC RX MEDLINE=94216346; PubMed=7512964;

```

```

RA Mei L., Doherty C.A., Haganir R.L.;
RA "RNA splicing regulates the activity of a SH2 domain-containing
RA protein tyrosine phosphatase."
RA J. Biol. Chem. 269:12254-12262(1994).
RA [3]
RA PARTIAL SEQUENCE.
RA MEDLINE=93011127; PubMed=1382993; Suzuki Y., Tsuki S.;
RA Hiraga A., Munakata H., Hata K.,
RA "Purification and characterization of a rat liver protein-tyrosine
RA phosphatase with sequence similarity to src-homology region 2."
RA Eur. J. Biochem. 209:195-206(1992).
RA [4]
RA PTNS1 BINDING.
RA MEDLINE=97215901; PubMed=9062191;
RA Khaitonenkov A., Chen Z., Sures I., Wang H., Schilling J.,
RA Ullrich A.;
RA "A family of proteins that inhibit signalling through tyrosine kinase
RA receptors."
RA Nature 386:181-186(1997).
RA -!- FUNCTION: This PTPase activity may directly link growth factor
RA receptors and other signaling proteins through protein-tyrosine
RA phosphorylation. The SH2 regions may interact with other cellular
RA components to modulate its own phosphatase activity against
RA interacting substrates.
RA -!- CATALYTIC ACTIVITY: Protein tyrosine phosphatase + H(2)O = protein
RA tyrosine + phosphate.
RA -!- SUBUNIT: Binds PTNS1.
RA -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
RA -!- PTM: Phosphorylated by tyrosine-protein kinases (By similarity).
RA -!- SIMILARITY: Contains 2 SH2 domains.
RA
RA This SWISS-PROT entry is copyrighted. It is produced through a collaboration
RA between the Swiss Institute of Bioinformatics and the EMBL Outstation -
RA the European Bioinformatics Institute. There are no restrictions on its
RA use by non-profit institutions as long as its content is in no way
RA modified and this statement is not removed. Usage by and for commercial
RA entities requires a license agreement (See http://www.isb-sib.ch/announce/
RA or send an email to license@isb-sib.ch).
RA
RA -----
RA EMBL; U09307; AAA20543.1; -
RA EMBL; U05963; AAA19133.1; -
RA PIR; A53593; A53593.
RA HSP; P35235; IYXA.
RA InterPro; IPR000980; SH2.
RA InterPro; IPR000387; TYR_phosphatase.
RA InterPro; IPR000242; Tyr_PP.
RA Pfam; PF00017; SH2; 2.
RA Pfam; PF00102; Y_phosphatase; 1.
RA PRINTS; PR00700; PRTYPHPTASE.
RA PRINTS; PR00401; SH2DOMAIN.
RA PRODOM; PD000093; SH2; 2.
RA SMART; SM00194; PTPC; 1.
RA SMART; SM00252; SH2; 2.
RA PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
RA PROSITE; PS00383; TYR_PHOSPHATASE_2; 1.
RA PROSITE; PS00055; TYR_PHOSPHATASE_ptp; 1.
RA PROSITE; PS00001; SH2; 2.
RA Hydrolase; SH2 domain; Repeat; Phosphorylation.
RA FT DOMAIN 6 102
RA SH2 1.
RA FT DOMAIN 112 216
RA SH2 2.
RA FT DOMAIN 276 517
RA PROTEIN-TYROSINE PHOSPHATASE.
RA FT ACT SITE 459 459
RA BY SIMILARITY.
RA FT CONFLICT 75 75
RA A -> P (IN REF. 1).
RA FT CONFLICT 407 407
RA G -> GOALL (IN REF. 2).
RA FT CONFLICT 547 547
RA Y -> S (IN REF. 2).
RA SQ SEQUENCE 593 AA; 68033 MW; 3329F10F0F60AF48 CRC64;
RA
RA Query Match 80.5%; Score 33; DB 1; Length 593;
RA Best Local Similarity 71.4%; Pred. No. 40;
RA Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
RA
RA QY 1 HYAKNPI 7
RA |||||

```

```
db      196 HYKNPM 202

RESULT 9
RRPL VSVJO STANDARD; PRT; 2109 AA.
ID ID RPRL VSVJO STANDARD; PRT; 2109 AA.
AC AC P16379;
DT DT 01-AUG-1990 (Rel. 15, Created)
DT DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DE RNA polymerase beta subunit (EC 2.7.7.48) (Large structural protein)
DE DE (L protein).
OS L.
OS Vesicular stomatitis virus (serotype New Jersey / strain Ogden).
OC OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC OC Rhabdoviridae; Vesiculovirus.
CC CC NCBI_TaxID=11283;
RN RN [1] SEQUENCE FROM N.A.
RP RP MEDLINE=90177235; PubMed=2155516;
RX RX Barik S., Rud E.W., Luk D., Banerjee A.K., Kang C.Y.;
XT XT "Nucleotide sequence analysis of the L gene of vesicular stomatitis
RT RT virus (New Jersey serotype): identification of conserved domains in L
RT RT proteins of nonsegmented negative-strand RNA viruses.";
RT RT virology 175:332-337(1990)
CC CC -!- FUNCTION: THIS PROTEIN IS PROBABLY A COMPONENT OF THE ACTIVE
CC CC POLYMERASE. IT MAY FUNCTION IN RNA SYNTHESIS, CAPPING, AS WELL AS
CC CC METHYLATION OF CAPS, AND POLY(A) SYNTHESIS.
CC CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC CC {RNA} (N).
CC CC -!- SUBUNIT: THOUGHT TO FORM A TRANSCRIPTION COMPLEX WITH THE
CC CC NUCLEOCAPSID (N) PROTEIN.
CC CC -!- SIMILARITY: WITH THE L PROTEIN OF OTHER RHABDOVIRUSES AND
CC CC PARAMYXOVIRUSES.
-----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
-----
CC CC EMBL; M29788; AAA48442.1; -.
CC CC FIP; A46309; A46309.
CC CC InterPro; IPR002877; FtsJ.
CC CC InterPro; IPR007098; RNA_pol_monon.
CC CC InterPro; IPR001016; Viral_RNA_pol_L.
CC CC Pfam; PF01728; FtsJ; 1.
CC CC Pfam; PF00946; Paramyx_RNA_pol; 1.
CC CC Transferase; RNA-directed RNA polymerase.
CC CC SEQUENCE 2109 AA; 242111 MW; 724CF90ECE26CAB9 CRC64;

Query Match 80.5%; Score 33; DB 1; Length 2109;
Best Local Similarity 71.4%; Pred.No. 1.4e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HVAKNDPI 7
|::|||
DB 824 HFAENPI 830

RESULT 10
RRPL VSVSJ
ID ID RPRL VSVSJ STANDARD; PRT; 2109 AA.
AC AC P03523;
DT DT 21-JUL-1986 (Rel. 01, Created)
DT DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DE RNA polymerase beta subunit (EC 2.7.7.48) (Large structural protein)
DE DE (L protein).
OS L.
OS Vesicular stomatitis virus (strain San Juan)
OC OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; AE000735; AAC07313.1; -  
DR F1R; D70416; D70416.  
DR HAMAP; MF 00495; -; 1.  
DR InterPro; IPR006402; HAD-SF-IA-V3.  
DR InterPro; IPR006439; HAD-SF-A-V1.  
DR InterPro; IPR005933; Hlgase/Hydrolase.  
DR InterPro; IPR005934; Hydrolase.  
DR Pfam; PF00702; Hydrolase; 1.  
DR PRINTS; PR00413; HADHALOGNASE.  
DR TIGRFAMs; TIGR01549; HAD-SF-IA-V1; 1.  
DR TIGRFAMs; TIGR01509; HAD-SF-IA-V3; 1.  
KW Carbohydrate metabolism; Hydrolase; Complete proteome.  
SQ SEQUENCE 213 AA; 23895 MW; 0722EC7F6B8E48753 CRC64;

Query Match 75.6%; Score 31; DB 1; Length 213;  
Best Local Similarity 57.1%; Pred. No. 37;  
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 HYAKNPI 7  
DB 71 HYLENPV 77

RESULT 12  
ID OPSE ANGAN STANDARD; PRT; 352 AA.  
AC Q90215;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Rhodopsin, freshwater eel.  
OS Anguilla anguilla (European freshwater eel).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Anguilliformes; Anguillidae;  
OC Anguilla.  
OX NCBI\_TaxID=7936;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Retina.  
RX MEDLINE=96156843; PubMed=8587887;  
RA Archer S., Hope A., Partridge J.C.;  
RT "The molecular basis for the green-blue sensitivity shift in the rod  
visual pigments of the European eel."  
RL Proc. R. Soc. Lond., B, Biol. Sci. 262:289-295(1995).  
CC -!- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT  
MEDIATE VISION. THEY CONSIST OF AN APOPROTEIN, OPSIN, COVALENTLY  
LINKED TO CIS-RETINAL.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -!- TISSUE SPECIFICITY: ROD SHAPED PHOTORECEPTOR CELLS WHICH MEDIATES  
VISION IN DIM LIGHT.  
CC -!- DEVELOPMENTAL STAGE: WHEN EEL MATURES SEXUALLY AND MIGRATES BACK  
TO DEEP SEA BREEDING GROUNDS THE VISUAL PIGMENTS IN ITS ROD  
PHOTORECEPTORS CHANGE FROM BEING MAXIMALLY SENSITIVE TO GREEN  
LIGHT TO BEING MAXIMALLY SENSITIVE TO BLUE LIGHT. IN PART, THIS  
CHANGE IN SENSITIVITY IS DUE TO A CHANGE IN THE OPSIN COMPONENT OF  
THE VISUAL PIGMENT MOLECULE. THIS GREEN SENSITIVE RHODOPSIN IS  
EXPRESSED DURING LIFE IN GREENER INLAND AND COASTAL WATERS.  
CC -!- PTM: SOME OR ALL OF THE CARBOXYL-TERMINAL SER OR THR RESIDUES MAY  
BE PHOSPHORYLATED.  
CC -!- MISCELLANEOUS: THIS OPSIN HAS AN ABSORPTION MAXIMUM AROUND 501-523  
NM.  
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.  
CC OPSIN SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC -----

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; L78007; AAA99200.1; -  
DR HSSP; P02699; IBOJ.  
DR InterPro; IPR000276; GPCR Rhodopsin.  
DR InterPro; IPR001760; Opsin.  
DR Pfam; PF00001; 7tm\_1; 1.  
DR PRINTS; PR00237; GPCRHOPOPSN.  
DR PROSITE; PS00237; G-PROTEIN\_RECEP\_F1\_1; 1.  
DR PROSITE; PS00262; G-PROTEIN\_RECEP\_F1\_2; 1.  
DR PROSITE; PS00238; OPSIN; 1.  
KW Photoreceptor; Retinal protein; Transmembrane; Glycoprotein; Vision;  
KW Phosphorylation; Lipoprotein; Palmitate; G-protein coupled receptor.  
FT DOMAIN 1 36  
FT TRANSMEM 37 61  
FT DOMAIN 62 73  
FT TRANSMEM 74 98  
FT DOMAIN 99 113  
FT TRANSMEM 114 133  
FT DOMAIN 134 152  
FT TRANSMEM 153 176  
FT DOMAIN 177 202  
FT TRANSMEM 203 230  
FT DOMAIN 231 252  
FT TRANSMEM 253 276  
FT DOMAIN 277 284  
FT TRANSMEM 285 309  
FT DOMAIN 310 352  
FT DOMAIN 329 332  
FT CARBOHYD 2 2  
FT CARBOHYD 15 15  
FT CARBOHYD 200 200  
FT BINDING 296 296  
FT LIPID 323 323  
FT DISULFID 110 187  
SQ SEQUENCE 352 AA; 39468 MW; 65578919D6679DB4 CRC64;

Query Match 75.6%; Score 31; DB 1; Length 352;  
Best Local Similarity 83.3%; Pred. No. 60;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HYAKNP 6  
DB 191 HYPNP 196

RESULT 13  
ID LEM1 MACMU STANDARD; PRT; 372 AA.  
AC Q95158;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE L-selectin precursor (lymph node homing receptor) (leukocyte adhesion  
molecule-1) (LAM-1) (Leukocyte-endothelial cell adhesion molecule 1)  
(LECAM1) (CD62L).  
GN SELL.  
OS Macaca mulatta (Rhesus macaque).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
OC Cercopithecoidea; Macaca.  
OX NCBI\_TaxID=9544;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Budman J.I., Fu H., Johnson C.E., Thakur A.B., Berg E.L.,  
RA Tsurushita N.;  
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: CELL SURFACE ADHESION PROTEIN. MEDIATES THE ADHERENCE  
OF LYMPHOCYTES TO ENDOTHELIAL CELLS OF HIGH ENDOTHELIAL



Search completed: February 11, 2004, 17:04:04  
Job time : 6.16667 sec

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 16:55:29 ; Search time 9.33333 Seconds  
(without alignments)  
72.127 Million cell updates/sec

Title: US-09-901-187C-2

Perfect score: 41

Sequence: 1 HYAKNPI 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168692 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 76:\*

1: Pirl:\*

2: Pirl:\*

3: Pirl:\*

4: Pirl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37	90.2	97	2 A86848	hypothetical prote
2	36	87.8	257	2 T06414	cytochrome-c oxida
3	36	87.8	258	1 OBPM2	cytochrome-c oxida
4	36	87.8	258	2 S13533	cytochrome-c oxida
5	36	87.8	260	2 S07169	cytochrome-c oxida
6	36	87.8	260	2 S14157	cytochrome-c oxida
7	36	87.8	260	2 S17300	cytochrome-c oxida
8	36	87.8	306	2 S14455	cytochrome-c oxida
9	36	87.8	354	4 A27071	cytoplasmic male s
10	35	85.4	989	2 B84532	hypothetical prote
11	34	82.9	342	2 T45227	probable F420H2-de
12	34	82.9	551	2 B64005	hypothetical prote
13	33	80.5	176	2 E72348	conserved hypothet
14	33	80.5	200	2 B64653	probable outer mem
15	33	80.5	252	2 S25950	cytochrome-c oxida
16	33	80.5	383	2 G83500	succinyl-diaminopi
17	33	80.5	465	1 T51095	acid phosphatase
18	33	80.5	465	2 T51094	acid phosphatase
19	33	80.5	522	2 AD2344	hypothetical prote
20	33	80.5	585	2 A6209	protein-tyrosine-p
21	33	80.5	593	1 JN0805	protein-tyrosine-p
22	33	80.5	593	2 JC5167	protein-tyrosine-p
23	33	80.5	595	1 A55651	protein-tyrosine-p
24	33	80.5	597	1 A53593	protein-tyrosine-p
25	33	80.5	827	2 S25949	gene coxII intron
26	33	80.5	989	2 I56333	apolipoprotein B -
27	33	80.5	2109	1 ZLVN	genome polypeptide
28	33	80.5	2109	1 A45309	genome polypeptide
29	32	78.0	176	2 D72688	hypothetical prote

30	32	78.0	258	2 T11923	cytochrome-c oxida
31	32	78.0	404	2 B81071	conserved hypothet
32	32	78.0	422	2 F81804	hypothetical prote
33	32	78.0	533	2 T13607	hypothetical prote
34	32	78.0	537	2 F86336	Fl4010.10 protein
35	32	78.0	560	2 T43188	heat shock protein
36	32	78.0	630	2 D97592	hypothetical prote
37	32	78.0	946	2 A71843	d-lactate dehydrog
38	32	78.0	1646	2 T20740	hypothetical prote
39	31	75.6	137	2 G72666	hypothetical prote
40	31	75.6	204	2 T47721	hypothetical prote
41	31	75.6	213	1 D70416	phosphoglycolate p
42	31	75.6	250	2 AE3607	lysosyme (SC 3.2.1
43	31	75.6	283	2 AI2836	glycosyl hydrolase
44	31	75.6	287	2 AI2870	cytochrome c oxida
45	31	75.6	316	2 G97452	cytochrome c oxida

## ALIGNMENTS

### RESULT 1

A86848  
hypothetical protein yseD [imported] - Lactococcus lactis subsp. lactis (strain IL1403)  
C:Species: Lactococcus lactis subsp. lactis  
C:Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001  
C:Accession: A86848  
R:Polotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Ehrh  
Genome Res. 11, 731-753, 2001  
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s  
A:Reference number: A86825; MUID:21235186; PMID:11337471  
A:Accession: A86848  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-97 <STO>  
A:Cross-references: DB:AB005176; PID:g12724809; PIDN:AAK05883.1; GSPDB:GN00146  
A:Experimental source: strain IL1403  
C:Genetics:  
A:Gene: yseD

Query Match 90.2%; Score 37; DB 2; Length 97;  
Best Local Similarity 85.7%; Pred. No. 1.6;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HYAKNPI 7  
|||  
DB 26 HYGNPI 32

### RESULT 2

T06414  
cytochrome-c oxidase (EC 1.9.3.1) chain II - garden pea mitochondrion (fragment)  
C:Species: mitochondrion Pisum sativum (garden pea)  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 21-Jul-2000  
C:Accession: T06414  
R:Covello, P.S.; Gray, M.W.  
Nucleic Acids Res. 18, 5189-5195, 1990  
A:Title: Differences in editing at homologous sites in messenger RNAs from angiosperm  
A:Reference number: Z15660; MUID:90384819; PMID:1698279  
A:Accession: T06414  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: mRNA  
A:Residues: 1-257 <COV>  
A:Cross-references: EMBL:X52866; NID:gl4318; PIDN:CAA37047.1; PID:g578925  
A:Experimental source: cv. Laxton's Progress  
C:Genetics:  
A:Gene: mitochondrion  
C:Function:  
A:Note: COX2  
A:Description: cytochrome-c oxidase complex catalyzes the oxidation of four molecules c  
rom the mitochondrial matrix producing two molecules of water and lowering the concentr  
A:Pathway: oxidative phosphorylation, respiratory chain  
C:Superfamily: cytochrome-c oxidase chain II; cytochrome-c oxidase chain II homology



C:Superfamily: cytochrome-c oxidase chain II; cytochrome-c oxidase chain II homology  
C:Keywords: copper; electron transfer; membrane-associated complex; mitochondrial inner  
F;24-240/Domain: cytochrome-c oxidase chain II homology <CO2>  
F;222,224,226,230/Binding site: copper 2 (Cys, Glu, Cys, His) #status predicted  
F;224/Binding site: magnesium (Glu) (shared with chain I) #status predicted

Query Match 87.8%; Score 36; DB 2; Length 258;  
Best Local Similarity 85.7%; Pred. No. 7.4;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HYAKNPI 7  
65 HYKKNPI 71

Db

RESULT 5

S07169  
cytochrome-c oxidase (EC 1.9.3.1) chain II - soybean mitochondrion  
C:Species: mitochondrion Glycine max (soybean)  
C:Date: 16-Sep-1992 #sequence\_revision 16-Sep-1992 #text\_change 20-Aug-1999  
C:Accession: S07169  
R;Grabau, E.A.  
Curr. Genet. 11, 287-293, 1987  
A:Title: Cytochrome oxidase subunit II gene is adjacent to an initiator methionine trans  
A:Reference number: S07169; MUID:88194682; PMID:2834094  
A:Accession: S07169  
A:Molecule type: DNA  
A:Residues: 1-260 <GRA>  
A:Cross-references: EMBL:X04825; NID:gl2974; PIDN:CAA28508.1; PID:gl2975  
A:Note: the authors translated the codon CGG for residues 54, 84, 126, and 247 as Trp.  
C:Comment: RNA editing in the maize mitochondrion results in a Met rather than a Thr at  
his translation.  
C:Genetics:  
A:Gene: COII  
A:Genome: mitochondrion  
C:Superfamily: cytochrome-c oxidase chain II; cytochrome-c oxidase chain II homology  
C:Keywords: copper; electron transfer; heme; membrane-associated complex; mitochondrial  
ein  
F;223-239/Domain: cytochrome-c oxidase chain II homology <CO2>  
F;186,221,225/Binding site: copper 1 (His, Cys, Cys) #status predicted  
F;221,223,225,229/Binding site: copper 2 (Cys, Glu, Cys, His) #status predicted  
F;223/Binding site: magnesium (Glu) (shared with chain I) #status predicted

Query Match 87.8%; Score 36; DB 2; Length 260;  
Best Local Similarity 85.7%; Pred. No. 7.5;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HYAKNPI 7  
64 HYKKNPI 70

Db

RESULT 6

S14157  
cytochrome-c oxidase (EC 1.9.3.1) chain II - sugar beet mitochondrion  
C:Species: mitochondrion Beta vulgaris var. altissima (sugar beet)  
C:Date: 16-Sep-1992 #sequence\_revision 16-Sep-1992 #text\_change 21-Jul-2000  
C:Accession: S14157  
R;Senda, M.; Harada, T.; Mikami, T.; Sugiyura, M.; Kinoshita, T.  
Curr. Genet. 19, 175-181, 1991  
A:Title: Genomic organization and sequence analysis of the cytochrome oxidase subunit I  
A:Reference number: S14138; MUID:91330331; PMID:1651175  
A:Accession: S14157  
A:Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-260 <SEN>  
A:Cross-references: GB:X57694; GB:S47710; GB:S47723; NID:gl1266; PIDN:CAA40875.1; PID:gl  
C:Genetics:  
A:Gene: coxII  
A:Genome: mitochondrion  
A:Introns: 128/1  
C:Superfamily: cytochrome-c oxidase chain II; cytochrome-c oxidase chain II homology  
C:Keywords: copper; electron transfer; heme; membrane-associated complex; mitochondria

ein  
F;224-240/Domain: cytochrome-c oxidase chain II homology <CO2>  
F;187,222,226,233/Binding site: copper 1 (His, Cys, Cys, Met) #status predicted  
F;222,224,226,230/Binding site: copper 2 (Cys, Glu, Cys, His) #status predicted  
F;224/Binding site: magnesium (Glu) (shared with chain I) #status predicted

Query Match 87.8%; Score 36; DB 2; Length 260;  
Best Local Similarity 85.7%; Pred. No. 7.5;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HYAKNPI 7  
|||  
Db 65 HYKNPI 71

RESULT 7  
S17300  
cytochrome-c oxidase (EC 1.9.3.1) chain II - beet mitochondrion  
C;Species: mitochondrion Beta vulgaris (beet)  
C;Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 20-Aug-1999  
C;Accession: S17300  
R;Mann, V.; Ekstein, I.; Nissen, H.; Hiser, C.; McIntosh, L.; Hirschberg, J.  
Plant Mol. Biol. 17, 559-566, 1991  
A;Title: The cytochrome oxidase II gene in mitochondria of the sugar-beet Beta vulgaris  
A;Reference number: S17300; MUID:91355954; PMID:1653062  
A;Accession: S17300  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-260 <MAN>  
A;Cross-references: EMBL:X55297  
C;Genetics:  
A;Genome: mitochondrion  
A;Introns: 128/1  
C;Superfamily: cytochrome-c oxidase chain II; cytochrome-c oxidase chain II homology  
C;Keywords: copper; electron transfer; heme; membrane-associated complex; mitochondrial  
ein

F;224-240/Domain: cytochrome-c oxidase chain II homology <CO2>  
F;187,222,226,233/Binding site: copper 1 (His, Cys, Cys, Met) #status predicted  
F;222,224,226,230/Binding site: copper 2 (Cys, Glu, Cys, His) #status predicted  
F;224/Binding site: magnesium (Glu) (shared with chain I) #status predicted

Query Match 87.8%; Score 36; DB 2; Length 260;  
Best Local Similarity 85.7%; Pred. No. 7.5;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HYAKNPI 7  
|||  
Db 65 HYKNPI 71

RESULT 8  
S14455  
cytochrome-c oxidase (EC 1.9.3.1) chain II - garden petunia mitochondrion  
C;Species: mitochondrion Petunia x hybrida (garden petunia)  
C;Date: 21-Nov-1993 #sequence\_revision 10-Nov-1995 #text\_change 20-Aug-1999  
C;Accession: S14455  
R;Fruitt, K.D.; Hanson, M.R.  
Curr. Genet. 16, 281-291, 1989  
A;Title: Cytochrome oxidase subunit II sequences in Petunia mitochondria: two intron-con  
A;Reference number: S14455; MUID:90182720; PMID:2560683  
A;Accession: S14455  
A;Molecule type: DNA  
A;Residues: 1-306 <PRU>  
A;Cross-references: EMBL:X17395  
A;Note: the authors translated the codon TTG for residue 57 as Trp and GAT for residue 1  
A;Note: the authors translated the codon CGG for residue 85 as Trp, CGG for residue 127  
C;Genetics:  
A;Genome: mitochondrion  
A;Introns: 128/3  
C;Superfamily: cytochrome-c oxidase chain II; cytochrome-c oxidase chain II homology  
C;Keywords: copper; electron transfer; membrane-associated complex; mitochondrial inner  
F;224-240/Domain: cytochrome-c oxidase chain II homology <CO2>  
F;222,224,226,230/Binding site: copper 2 (Cys, Glu, Cys, His) #status predicted

F;224/Binding site: magnesium (Glu) (shared with chain I) #status predicted

Query Match 87.8%; Score 36; DB 2; Length 306;  
Best Local Similarity 85.7%; Pred. No. 8.9;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HYAKNPI 7  
|||  
Db 65 HYKNPI 71

RESULT 9  
A27071  
cytoplasmic male sterility-associated mutant fusion protein - garden petunia mitochondrion  
N;Alternate names: CMS-associated protein  
C;Species: mitochondrion Petunia x hybrida (garden petunia)  
C;Date: 19-Nov-1988 #sequence\_revision 24-Oct-1996 #text\_change 24-Oct-1996  
C;Accession: A27071  
R;Young, E.G.; Hanson, M.R.  
Cell 50, 41-49, 1987  
A;Title: A fused mitochondrial gene associated with cytoplasmic male sterility is devel  
A;Reference number: A27071; MUID:87244322; PMID:2885095  
A;Accession: A27071  
A;Molecule type: DNA  
A;Residues: 1-354 <YOU>  
A;Cross-references: EMBL:M16770; NID:91256946  
A;Note: the authors translated the codon CTT for residue 94 as Pro, GGA for residue 224  
A;Note: the authors confirm that mRNA editing in plant mitochondria changes the Arg cod  
C;Comment: This sequence is the chimeric product of a fusion mutation.  
C;Genetics:  
A;Genome: mitochondrion  
C;Keywords: fusion protein; mitochondrion; tandem repeat  
F;1-35/Region: H-transferring ATP synthase lipid-binding protein  
F;40-186/Region: cytochrome-c oxidase chain II  
F;166-195/Region: 10-residue repeats (P-S-L-G-V-K-G-D-A-V)  
F;243-314/Region: 18-residue repeats (P-E-G-[ND]-P-Q-A-L-P-E-D-G-Q-F-H-A-[VI]-A)

Query Match 87.8%; Score 36; DB 4; Length 354;  
Best Local Similarity 85.7%; Pred. No. 10;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HYAKNPI 7  
|||  
Db 70 HYKNPI 76

RESULT 10  
B84532  
hypothetical protein At2g15690 [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C;Accession: B84532  
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,  
Nature 402, 761-768, 1999  
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A;Reference number: A84420; MUID:20083487; PMID:10617197  
A;Accession: B84532  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-989 <STO>  
A;Cross-references: GB:AE002093; NID:94335735; PIDN:AAD17413.1; GSPDB:GN00139  
C;Genetics:  
A;Gene: At2g15690  
A;Map position: 2

Query Match 85.4%; Score 35; DB 2; Length 989;  
Best Local Similarity 85.7%; Pred. No. 49;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HYAKNPI 7  
|||

Db 244 HYAKNPI 250

RESULT 11  
T45227  
N:Alternate names: f4db protein  
C:Species: Methanolobus tindarius  
C>Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 31-Mar-2000  
C:Accession: T45227  
R:Westenberg, D.J.; Braune, A.; Ruppert, C.; Mueller, V.; Herzberg, C.; Gottschalk, G.;  
submitted to the EMBL Data Library, September 1998  
A:Description: The P40H2-dehydrogenase from Methanolobus tindarius: Cloning of the ffd  
A:Reference number: Z22947  
A:Accession: T45227  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-342 <WES>  
A:Cross-references: EMBL:AJ011519; PIDN:CAB56640.1  
A:Experimental source: DSM 2278  
C:Genetics:  
A:Gene: ffdB  
C:Superfamily: unassigned ferredoxin 2 [4Fe-4S]-related proteins; ferredoxin 2 [4Fe-4S] h  
Query Match 82.9%; Score 34; DB 2; Length 342;  
Best Local Similarity 83.3%; Pred. No. 26;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 HYAKNP 6  
DB 248 HYARNP 253

RESULT 12  
B64005  
Hypothetical protein H10275 - Haemophilus influenzae (strain Rd KW20)  
C:Species: Haemophilus influenzae  
C>Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 08-Oct-1999  
C:Accession: B64005  
R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.  
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.  
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M.  
Science 269, 496-512, 1995  
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,  
A.; Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
A:Reference number: A64000; MUID:95350630; PMID:7542800  
A:Accession: B64005  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-551 <TIGR>  
A:Cross-references: GB:U32714; GB:I42023; NID:G1573241; PIDN:AA021949.1; PID:G1573251; T  
Query Match 82.9%; Score 34; DB 2; Length 551;  
Best Local Similarity 83.3%; Pred. No. 43;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 HYAKNP 6  
DB 145 HYSKNP 150

RESULT 13  
E72348  
conserved hypothetical protein - Thermotoga maritima (strain MSB8)  
C:Species: Thermotoga maritima  
C>Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C:Accession: E72348  
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey  
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;  
C.M.  
Nature 399, 323-329, 1999  
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq  
A:Reference number: A72200; MUID:99287316; PMID:10360571

A:Accession: E72348  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-176 <ARN>  
A:Cross-references: GB:AE001739; GB:AE000512; NID:G4981176; PIDN:AA035740.1; PID:G49811  
A:Experimental source: strain MSB8  
C:Genetics:  
A:Gene: TM0656

Query Match 80.5%; Score 33; DB 2; Length 176;  
Best Local Similarity 83.3%; Pred. No. 21;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 HYAKNP 6  
DB 158 HYSKNP 163

RESULT 14  
B64653  
probable outer membrane protein - Helicobacter pylori  
C:Species: Helicobacter pylori  
A:Variety: strains J99, 26695  
C>Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 29-Sep-1999  
C:Accession: B64653; G71941  
R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.  
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalaf, H.G.; Glodek, A.; McKen  
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L  
Nature 388, 539-547, 1997  
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C  
A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.  
A:Reference number: A64520; MUID:97394467; PMID:9252485  
A:Accession: B64653  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-200 <TOM>  
A:Cross-references: GB:AE000613; GB:AE000511; NID:G2314200; PIDN:AA08104.1; PID:G23142  
A:Experimental source: strain 26695  
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.  
; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;  
Nature 397, 178-180, 1999  
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric pat  
A:Reference number: A71800; MUID:99120557; PMID:9923682  
A:Accession: G71941  
A:Molecule type: DNA  
A:Residues: 1-200 <ARN>  
A:Cross-references: GB:AE001471; GB:AE001439; NID:G4154880; PIDN:AA05945.1; PID:G41548  
A:Experimental source: strain J99  
C:Genetics:  
A:Gene: jhp0359; HP1066  
A:Start codon: GTG  
C:Superfamily: hypothetical protein HP1066

Query Match 80.5%; Score 33; DB 2; Length 200;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 YAKNPI 7  
DB 101 YAKNPI 106

RESULT 15  
S25950  
cytochrome-c oxidase (EC 1.9.3.1) chain II - liverwort (Marchantia polymorpha) mitochon  
C:Species: Marchantia polymorpha  
C>Date: 07-May-1993 #sequence\_revision 07-May-1993 #text\_change 20-Aug-1999  
C:Accession: S25950  
R:Oda, K.; Yamato, K.; Ohta, E.; Nakamura, Y.; Takemura, M.; Nozato, N.; Akashi, K.; Ka  
J. Mol. Biol. 223, 1-7, 1992  
A:Title: Gene organization deduced from the complete sequence of liverwort Marchantia p  
A:Reference number: S25941; MUID:92114051; PMID:1731062  
A:Accession: S25950

A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-252 <ODA>  
A;Cross-references: EMBL:M68929  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1992  
C;Genetics:  
A;Gene: coxII  
A;Genome: mitochondrion  
A;Introns: 33/1; 84/1  
C;Superfamily: cytochrome-c oxidase chain II; cytochrome-c oxidase chain II homology  
C;Keywords: copper; electron transfer; heme; membrane-associated complex; mitochondrial  
ein  
F;22-238/Domain: cytochrome-c oxidase chain II homology <CO2>  
F;185,220,224,231/Binding site: copper 1 (His, Cys, Cys, Met) #status predicted  
F;220,222,224,228/Binding site: copper 2 (Cys, Glu, Cys, His) #status predicted  
F;222/Binding site: magnesium (Glu) (shared with chain I) #status predicted

Query Match 80.5%; Score 33; DB 2; Length 252;  
Best Local Similarity 71.4%; Pred. No. 30;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 HYAKNPI 7  
||| :|||  
Db 63 HYKSNPI 69

Search completed: February 11, 2004, 17:11:32  
Job time : 10.3333 secs



```

; Patent No. US20020160456A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean Francois
; APPLICANT: Omen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding No. US20020160456A1 Helicobacter Polypeptides in the
; TITLE OF INVENTION: Genome
; FILE REFERENCE: 06132/043002
; CURRENT APPLICATION NUMBER: US/09/895,913A
; CURRENT FILING DATE: 2001-06-29
; PRIOR FILING DATE: 1997-06-24
; NUMBER OF SEQ ID NOS: 368
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 132
; LENGTH: 190
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-895-913A-132

Query Match      80.5%; Score 33; DB 10; Length 190;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 YAKNPI 7
      |||||
Db      91 YAKNPI 96

RESULT 3
US-09-920-021A-3
; Sequence 3, Application US/09920021A
; Patent No. US20020110800A1
; GENERAL INFORMATION:
; APPLICANT: KAPLAN, DAVID
; APPLICANT: MARSH, H. NICK
; TITLE OF INVENTION: USE OF SHP-1 AND SHP-2 TO DETECT
; TITLE OF INVENTION: COMPOUNDS INVOLVED IN NEURONAL SURVIVAL
; FILE REFERENCE: 08338/016001
; CURRENT APPLICATION NUMBER: US/09/920,021A
; CURRENT FILING DATE: 2001-08-01
; PRIOR APPLICATION NUMBER: US/08/918,157
; PRIOR FILING DATE: 1997-08-25
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 593
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-920-021A-3

Query Match      80.5%; Score 33; DB 10; Length 593;
Best Local Similarity 71.4%; Pred. No. 2,6e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 HYAKNPI 7
      |||||
Db      196 HYAKNPM 202

RESULT 4
US-10-366-547-16
; Sequence 16, Application US/10366547
; Publication No. US2003021589A1
; GENERAL INFORMATION:
; APPLICANT: Meng, Tzu-Ching
; APPLICANT: Tonks, Nicholas K.
; APPLICANT: Cool, Deborah E.
; TITLE OF INVENTION: REVERSIBLE OXIDATION OF PROTEIN TYROSINE
; TITLE OF INVENTION: PHOSPHATASES

```

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 HYAKNPI 7  
Db 196 HYKKNPM 202

## RESULT 7

US-10-366-547-30  
; Sequence 30, Application US/10366547  
; Publication No. US20030215899A1  
; GENERAL INFORMATION:  
; APPLICANT: Meng, Tzu-Ching  
; APPLICANT: Tonks, Nicholas K.  
; APPLICANT: Cool, Deborah E.  
; TITLE OF INVENTION: REVERSIBLE OXIDATION OF PROTEIN TYROSINE  
; TITLE OF INVENTION: PHOSPHATASES  
; FILE REFERENCE: 200125.439  
; CURRENT APPLICATION NUMBER: US/10/366,547  
; CURRENT FILING DATE: 2003-02-12  
; NUMBER OF SEQ ID NOS: 98  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 30  
; LENGTH: 593  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-366-547-30

Query Match 80.5%; Score 33; DB 12; Length 593;  
Best Local Similarity 71.4%; Pred. No. 2.6e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 HYAKNPI 7  
Db 196 HYKKNPM 202

## RESULT 8

US-10-366-547-32  
; Sequence 32, Application US/10366547  
; Publication No. US20030215899A1  
; GENERAL INFORMATION:  
; APPLICANT: Meng, Tzu-Ching  
; APPLICANT: Tonks, Nicholas K.  
; APPLICANT: Cool, Deborah E.  
; TITLE OF INVENTION: REVERSIBLE OXIDATION OF PROTEIN TYROSINE  
; TITLE OF INVENTION: PHOSPHATASES  
; FILE REFERENCE: 200125.439  
; CURRENT APPLICATION NUMBER: US/10/366,547  
; CURRENT FILING DATE: 2003-02-12  
; NUMBER OF SEQ ID NOS: 98  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 32  
; LENGTH: 593  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-366-547-32

Query Match 80.5%; Score 33; DB 12; Length 593;  
Best Local Similarity 71.4%; Pred. No. 2.6e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 HYAKNPI 7  
Db 196 HYKKNPM 202

## RESULT 9

US-10-262-552-2  
; Sequence 2, Application US/10262552  
; Publication No. US20030125289A1  
; GENERAL INFORMATION:  
; APPLICANT: Gelb, Bruce D.

APPLICANT: Tartaglia, Marco  
; TITLE OF INVENTION: NOONAN SYNDROME GENE  
; FILE REFERENCE: 2420/1J859-US1  
; CURRENT APPLICATION NUMBER: US/10/262,552  
; CURRENT FILING DATE: 2002-10-01  
; PRIOR APPLICATION NUMBER: US 60/326,532  
; PRIOR FILING DATE: 2001-10-01  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 593  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-262-552-2

Query Match 80.5%; Score 33; DB 15; Length 593;  
Best Local Similarity 71.4%; Pred. No. 2.6e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 HYAKNPI 7  
Db 196 HYKKNPM 202

## RESULT 10

US-10-366-547-14  
; Sequence 14, Application US/10366547  
; Publication No. US20030215899A1  
; GENERAL INFORMATION:  
; APPLICANT: Meng, Tzu-Ching  
; APPLICANT: Tonks, Nicholas K.  
; APPLICANT: Cool, Deborah E.  
; TITLE OF INVENTION: REVERSIBLE OXIDATION OF PROTEIN TYROSINE  
; TITLE OF INVENTION: PHOSPHATASES  
; FILE REFERENCE: 200125.439  
; CURRENT APPLICATION NUMBER: US/10/366,547  
; CURRENT FILING DATE: 2003-02-12  
; NUMBER OF SEQ ID NOS: 98  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14  
; LENGTH: 597  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-10-366-547-14

Query Match 80.5%; Score 33; DB 12; Length 597;  
Best Local Similarity 71.4%; Pred. No. 2.6e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 HYAKNPI 7  
Db 196 HYKKNPM 202

## RESULT 11

US-10-038-010-22  
; Sequence 22, Application US/10038010  
; Publication No. US20030040089A1  
; GENERAL INFORMATION:  
; APPLICANT: HYBRIGENICS  
; APPLICANT: Pierre, Legrain  
; TITLE OF INVENTION: Protein-protein interactions in adipocyte cells  
; FILE REFERENCE: B4767A  
; CURRENT APPLICATION NUMBER: US/10/038,010  
; CURRENT FILING DATE: 2002-07-23  
; PRIOR APPLICATION NUMBER: US 60/259,377  
; PRIOR FILING DATE: 2001-01-02  
; NUMBER OF SEQ ID NOS: 67  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 22  
; LENGTH: 597  
; TYPE: PRT  
; ORGANISM: Homo sapiens

```
;
; FEATURE:
; NAME/KEY: hshp2.fl
; LOCATION: (1), (597)
; OTHER INFORMATION:
US-10-038-010-22

Query Match      80.5%; Score 33; DB 15; Length 597;
Best Local Similarity 71.4%; Pred. No. 2.6e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 HYAKNP 7
      |||||
Db      196 HYKNPM 202

RESULT 12
US-10-156-761-10729
; Sequence 10729, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 10729
; LENGTH: 701
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-10729

Query Match      80.5%; Score 33; DB 15; Length 701;
Best Local Similarity 83.3%; Pred. No. 3.1e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 HYAKNP 6
      |||||
Db      381 HYKNPM 386

RESULT 13
US-10-084-846A-110
; Sequence 110, Application US/10084846A
; Publication No. US2004006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLENWEG, AGNES
; APPLICANT: TREPFER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084,846A
; CURRENT FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 110
; LENGTH: 790
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-891-3943
; Sequence 3943, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3943
; LENGTH: 31
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-891-3943
Query Match      78.0%; Score 32; DB 11; Length 31;
Best Local Similarity 83.3%; Pred. No. 21;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

;
; ORGANISM: Streptomyces viridochromogenes
; FEATURE:
; OTHER INFORMATION: amino acid sequence encoded by aviX14 dna
US-10-084-846A-110

Query Match      80.5%; Score 33; DB 12; Length 790;
Best Local Similarity 83.3%; Pred. No. 3.5e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 HYAKNP 6
      |||||
Db      413 HYKNPM 418

RESULT 14
US-10-084-846A-3
; Sequence 3, Application US/10084846A
; Publication No. US2004006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLENWEG, AGNES
; APPLICANT: TREPFER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084,846A
; CURRENT FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 3
; LENGTH: 19695
; TYPE: PRT
; ORGANISM: Streptomyces viridochromogenes
; FEATURE:
; OTHER INFORMATION: Protein 1: amino acid sequence encoded by coding strand 1.
; OTHER INFORMATION: Start codon: gga, Start position: nucleotide 1.
US-10-084-846A-3

Query Match      80.5%; Score 33; DB 12; Length 19695;
Best Local Similarity 83.3%; Pred. No. 8.7e+03;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 HYAKNP 6
      |||||
Db      17568 HYKNPM 17573

RESULT 15
US-09-764-891-3943
; Sequence 3943, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3943
; LENGTH: 31
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-891-3943
Query Match      78.0%; Score 32; DB 11; Length 31;
Best Local Similarity 83.3%; Pred. No. 21;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```



Oy 1 HYAKP 6  
||  
Db 23 HYKRP 28

Search completed: February 11, 2004, 17:54:04  
Job time : 25.6667 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 16:56:09 ; Search time 10.4167 Seconds  
(without alignments)  
28.433 Million cell updates/sec

Title: US-09-901-187C-2

Perfect score: 41

Sequence: 1 HYAKNPI 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*

- 1: /cgn2\_6/prodata/1/iaa/5A-COMB.pep.\*
- 2: /cgn2\_6/prodata/1/iaa/5B-COMB.pep.\*
- 3: /cgn2\_6/prodata/1/iaa/6A-COMB.pep.\*
- 4: /cgn2\_6/prodata/1/iaa/6B-COMB.pep.\*
- 5: /cgn2\_6/prodata/1/iaa/PCTUS-COMB.pep.\*
- 6: /cgn2\_6/prodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	* Query Match	Length	DB ID	Description
1	33	80.5	102	1	US-08-202-389-27
2	33	80.5	290	4	US-09-252-991A-19355
3	33	80.5	483	3	US-08-369-822C-22
4	33	80.5	483	3	US-08-582-776C-37
5	33	80.5	483	3	US-08-434-831B-34
6	33	80.5	593	1	US-08-202-389-12
7	33	80.5	593	1	US-08-018-129-5
8	33	80.5	593	2	US-08-448-250-5
9	33	80.5	593	4	US-09-282-257-5
10	33	80.5	676	4	US-09-107-532A-3847
11	33	80.5	1627	4	US-09-328-352-6604
12	33	80.5	2109	3	US-08-646-695-6
13	33	80.5	2109	5	PCT-US96-06053-6
14	32	78.0	293	4	US-09-328-352-4692
15	32	78.0	560	4	US-08-928-692-53
16	32	78.0	560	4	US-09-339-972-53
17	31	75.6	262	4	US-09-328-352-5407
18	31	75.6	279	2	US-08-326-286-7
19	31	75.6	490	4	US-09-107-532A-5048
20	31	75.6	599	2	US-08-846-526-11
21	31	75.6	599	3	US-08-172-339-2
22	31	75.6	599	4	US-09-398-395A-22
23	31	75.6	599	4	US-09-887-586A-22
24	31	75.6	599	4	US-08-895-752-22
25	31	75.6	599	4	US-09-903-012B-22
26	30	73.2	92	1	US-08-202-389-28
27	30	73.2	292	4	US-09-501-115-28

28	30	73.2	398	4	US-09-134-001C-4353	Sequence 4353, Ap
29	30	73.2	406	4	US-09-328-352-5309	Sequence 5309, Ap
30	30	73.2	3200	2	US-08-477-451-8	Sequence 8, Appli
31	29	70.7	18	4	US-09-292-225-12	Sequence 12, Appli
32	29	70.7	55	3	US-09-019-095A-32	Sequence 32, Appli
33	29	70.7	119	1	US-08-340-539A-14	Sequence 14, Appli
34	29	70.7	217	2	US-08-874-832-1	Sequence 1, Appli
35	29	70.7	217	2	US-08-874-832-2	Sequence 2, Appli
36	29	70.7	217	2	US-08-874-832-3	Sequence 3, Appli
37	29	70.7	217	2	US-08-874-832-4	Sequence 4, Appli
38	29	70.7	217	3	US-09-097-233-1	Sequence 1, Appli
39	29	70.7	217	3	US-09-097-233-2	Sequence 2, Appli
40	29	70.7	217	3	US-09-097-233-3	Sequence 3, Appli
41	29	70.7	217	3	US-09-097-233-4	Sequence 4, Appli
42	29	70.7	220	4	US-09-198-452A-461	Sequence 461, App
43	29	70.7	224	2	US-08-924-759-2	Sequence 2, Appli
44	29	70.7	224	3	US-09-248-335-2	Sequence 2, Appli
45	29	70.7	224	5	PCT-US91-09055-4	Sequence 4, Appli

## ALIGNMENTS

RESULT 1  
US-08-202-389-27  
; Sequence 27, Application US/08202389  
; Patent No. 5536636  
; GENERAL INFORMATION:  
; APPLICANT: Freeman Jr., Robert M.  
; APPLICANT: Plutzky, Jorge  
; APPLICANT: Neel, Benjamin G.  
; APPLICANT: Rosenberg, Robert D.  
; TITLE OF INVENTION: IDENTIFICATION OF NOVEL TYROSINE  
; TITLE OF INVENTION: PHOSPHATASES HAVING SH2 DOMAINS  
; NUMBER OF SEQUENCES: 54  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
; STREET: Two Militia Drive  
; CITY: Lexington  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02173  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/202.389  
; FILING DATE: 28-FEB-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/983,926  
; FILING DATE: 01-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/829,141  
; FILING DATE: 31-JAN-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/721,112  
; FILING DATE: 26-JUN-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Granahan, Patricia  
; REGISTRATION NUMBER: 32,227  
; REFERENCE/DOCKET NUMBER: BIH92-05WA  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 861-6240  
; TELEFAX: (617) 861-9540  
; INFORMATION FOR SEQ ID NO: 27:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 102 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein

US-08-202-389-27

Query Match 80.5%; Score 33; DB 1; Length 102;  
Best Local Similarity 71.4%; Pred. No. 10;  
Matches 5; Conservative 1; Mismatches 0; Gaps 0;

Qy 1 HYAKNPI 7  
|||  
Db 85 HYKKNPM 91

RESULT 2

US-09-252-991A-19355  
; Sequence 19355, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 19355  
; LENGTH: 290  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-19355

Query Match 80.5%; Score 33; DB 4; Length 290;  
Best Local Similarity 85.7%; Pred. No. 32;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 HYAKNPI 7  
|||  
Db 113 HIAKNPI 119

RESULT 3

US-08-369-822C-22  
; Sequence 22, Application US/08369822C  
; Patent No. 6015860  
; GENERAL INFORMATION:  
; APPLICANT: Lipkin, W. I.  
; APPLICANT: Briese, Thomas  
; APPLICANT: Kliche, Stefanie  
; APPLICANT: Schneider, Patrick A.  
; APPLICANT: Stitz, Lothar  
; APPLICANT: Schneemann, Anette  
; TITLE OF INVENTION: Borna Disease Viral Sequences,  
; TITLE OF INVENTION: Diagnostics and Therapeutics for Central Nervous  
; TITLE OF INVENTION: System Diseases  
; NUMBER OF SEQUENCES: 46  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fulbright & Jaworski, L.L.P.  
; STREET: 865 South Figueroa Street, 29th Floor  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: USA  
; ZIP: 90017-2571  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: WINDOWS NT-WORDEPERFECT 8.0  
; SOFTWARE: ASCII (DOS) TEXT  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/369,822C  
; FILING DATE: 06-JAN-1995  
; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:  
; NAME: Churchill, Margaret A. (Ph.D.)  
; REGISTRATION NUMBER: 39,944  
; REFERENCE/DOCKET NUMBER: 1279-194XX  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 213/892-9200  
; TELEFAX: 213/680-4518  
; INFORMATION FOR SEQ ID NO: 22:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 483 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHEICAL: NO  
; ANTI-SENSE: NO  
US-08-369-822C-22

Query Match 80.5%; Score 33; DB 3; Length 483;  
Best Local Similarity 71.4%; Pred. No. 55;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HYAKNPI 7  
|||  
Db 395 HFAENPI 401

RESULT 4

US-08-582-776C-37  
; Sequence 37, Application US/08582776C  
; Patent No. 6077510  
; GENERAL INFORMATION:  
; APPLICANT: Lipkin, W. I.  
; APPLICANT: Briese, Thomas  
; APPLICANT: Kliche, Stefanie  
; APPLICANT: Schneider, Patrick A.  
; APPLICANT: Stitz, Lothar  
; APPLICANT: Schneemann, Anette  
; TITLE OF INVENTION: Borna Disease Viral Sequences,  
; TITLE OF INVENTION: Diagnostics and Therapeutics for Central Nervous  
; TITLE OF INVENTION: System Diseases  
; NUMBER OF SEQUENCES: 61  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fulbright & Jaworski, L.L.P.  
; STREET: 865 South Figueroa Street, 29th Floor  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: USA  
; ZIP: 90017-2576  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: WINDOWS NT  
; SOFTWARE: ASCII DOS TEXT  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/582,776C  
; FILING DATE: 04-JAN-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/369,822  
; FILING DATE: 06-JAN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/434,831  
; FILING DATE: 04-MAY-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Churchill, Margaret A.  
; REGISTRATION NUMBER: 39,944  
; REFERENCE/DOCKET NUMBER: 1279-194C2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 213/892-9200  
; TELEFAX: 213/680-4518  
; INFORMATION FOR SEQ ID NO: 37:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 483 amino acids

Db 395 HFAENPI 401

RESULT 6

US-08-202-389-12

Sequence 12, Application US/08202389

Patent No. 553636

GENERAL INFORMATION:

APPLICANT: Freeman Jr., Robert M.

APPLICANT: Plutsky, Jorge

APPLICANT: Neel, Benjamin G.

APPLICANT: Rosenberg, Robert D.

TITLE OF INVENTION: IDENTIFICATION OF NOVEL TYROSINE

TITLE OF INVENTION: PHOSPHATASES HAVING SH2 DOMAINS

NUMBER OF SEQUENCES: 54

CORRESPONDENCE ADDRESS:

ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.

STREET: Two Militia Drive

CITY: Lexington

STATE: MA

COUNTRY: USA

ZIP: 02173

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/202,389

FILING DATE: 28-FEB-1994

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/983,926

FILING DATE: 01-DEC-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/829,141

FILING DATE: 31-JAN-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/721,112

FILING DATE: 26-JUN-1991

ATTORNEY/AGENT INFORMATION:

NAME: Granahan, Patricia

REGISTRATION NUMBER: 32,227

REFERENCE/DOCKET NUMBER: BIH92-05MA

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 861-6240

TELEFAX: (617) 861-9540

INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:

LENGTH: 593 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-202-389-12

Query Match 80.5%; Score 33; DB 1; Length 593;

Best Local Similarity 71.4%; Pred. No. 69;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 HYAKNPI 7

Db 196 HYCKNPM 202

RESULT 7

US-08-018-129-5

Sequence 5, Application US/08018129

Patent No. 5589375

GENERAL INFORMATION:

APPLICANT: Ullrich, Axel

APPLICANT: Vogel, Wolfgang

TITLE OF INVENTION: PTP ID: A NOVEL PROTEIN TYROSINE

TITLE OF INVENTION: PHOSPHATASE

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

HYPOTHEICAL: NO

ANTI-SENSE: NO

US-08-582-776C-37

Query Match 80.5%; Score 33; DB 3; Length 483;

Best Local Similarity 71.4%; Pred. No. 55;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 HYAKNPI 7

Db 395 HFAENPI 401

RESULT 5

US-08-434-831B-34

Sequence 34, Application US/0843481B

Patent No. 6113905

GENERAL INFORMATION:

APPLICANT: Lipkin, W. I.

APPLICANT: Briese, Thomas

APPLICANT: Kliche, Stefanie

APPLICANT: Schneider, Patrick A.

APPLICANT: Stitz, Lothar

APPLICANT: Schneemann, Anette

TITLE OF INVENTION: Borna Disease Viral Sequences,

TITLE OF INVENTION: Diagnostics and Therapeutics for Central Nervous

TITLE OF INVENTION: System Diseases

NUMBER OF SEQUENCES: 59

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fulbright & Jaworski, L.L.P.

STREET: 865 South Figueroa Street, 29th Floor

CITY: Los Angeles

STATE: California

COUNTRY: USA

ZIP: 90017-2571

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: WINDOWS NT WORDPERFECT 8.0

SOFTWARE: ASCII (DOS) TEXT

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/434,831B

FILING DATE: 04-MAY-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/369,822

FILING DATE: 06-JAN-1995

ATTORNEY/AGENT INFORMATION:

NAME: Churchill, Margaret A. (Ph.D.)

REGISTRATION NUMBER: 39,944

REFERENCE/DOCKET NUMBER: 1279-194C1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 213/892-9200

TELEFAX: 213/680-4518

INFORMATION FOR SEQ ID NO: 34:

SEQUENCE CHARACTERISTICS:

LENGTH: 483 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

HYPOTHEICAL: NO

ANTI-SENSE: NO

US-08-434-831B-34

Query Match 80.5%; Score 33; DB 3; Length 483;

Best Local Similarity 71.4%; Pred. No. 55;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 HYAKNPI 7

NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS  
STREET: 1155 Avenue of Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/018,129  
FILING DATE: 19930216  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mirock, S. Leslie  
REGISTRATION NUMBER: 18,672  
REFERENCE/DOCKET NUMBER: 7683-017  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 593 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-018-129-5

```
Query Match      80.5%; Score 33; DB 1; Length 593;
Best Local Similarity 71.4%; Pred. No. 69;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

Qy 1 HYAKNPI 7  
|||  
db 196 HYKKNEM 202

RESULT 8  
US-08-448-250-5  
; Sequence 5, Application US/08448250  
; Patent No. 5981251  
; GENERAL INFORMATION:  
; APPLICANT: Ullrich, Axel  
; APPLICANT: Vogel, Wolfgang  
; TITLE OF INVENTION: PTP LD: A NOVEL PROTEIN TYROSINE  
; TITLE OF INVENTION: PHOSPHATASE  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS  
; STREET: 1155 Avenue of Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; \* COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/448,250  
; FILING DATE: 23-MAY-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/018,129  
; FILING DATE: 16-FEB-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Misrock, S Leslie

; CURRENT APPLICATION NUMBER: US/09/328,352  
 ; CURRENT FILING DATE: 1999-06-04  
 ; NUMBER OF SEQ ID NOS: 8252  
 ; SEQ ID NO 6604  
 ; LENGTH: 1627  
 ; TYPE: PRT  
 ; ORGANISM: Acinetobacter baumannii  
 US-09-328-352-6604  
 Query Match 80.5%; Score 33; DB 4; Length 1627;  
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 YAKNPI 7  
 DB 1487 YAKNPI 1492  
 RESULT 12  
 US-08-646-695-6  
 ; Sequence 6, Application US/08646695  
 ; Patent No. 6168943  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rose, John K.  
 ; TITLE OF INVENTION: RECOMBINANT VESICULOVIRUSES AND THEIR  
 ; TITLE OF INVENTION: USES  
 ; NUMBER OF SEQUENCES: 44  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: PENNIE & EDMONDS  
 ; STREET: 1155 Avenue of the Americas  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: USA  
 ; ZIP: 10036-2711  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent in Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/646,695  
 ; FILING DATE: On Even Date Herewith  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Mistrock, S. Leslie  
 ; REGISTRATION NUMBER: 18,872  
 ; REFERENCE/DOCKET NUMBER: 6523-008  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (212) 790-9090  
 ; TELEFAX: (212) 869-9741/8864  
 ; TELEX: 66141 PENNIE  
 ; INFORMATION FOR SEQ ID NO: 6:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2109 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 US-08-646-695-6  
 Query Match 80.5%; Score 33; DB 3; Length 2109;  
 Best Local Similarity 71.4%; Pred. No. 2.8e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 HYAKNPI 7  
 DB 824 HFAENPI 830  
 RESULT 13  
 PCT-US96-06053-6  
 ; Sequence 6, Application PC/TUS9606053  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Yale University

; CURRENT APPLICATION NUMBER: US/09/107,532A  
 ; CURRENT FILING DATE: 2000-06-04  
 ; NUMBER OF SEQ ID NOS: 8252  
 ; SEQ ID NO 6604  
 ; LENGTH: 1627  
 ; TYPE: PRT  
 ; ORGANISM: Acinetobacter baumannii  
 US-09-328-352-6604  
 Query Match 80.5%; Score 33; DB 4; Length 1627;  
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 YAKNPI 7  
 DB 1487 YAKNPI 1492  
 RESULT 12  
 US-08-646-695-6  
 ; Sequence 6, Application US/08646695  
 ; Patent No. 6168943  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rose, John K.  
 ; TITLE OF INVENTION: RECOMBINANT VESICULOVIRUSES AND THEIR  
 ; TITLE OF INVENTION: USES  
 ; NUMBER OF SEQUENCES: 44  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: PENNIE & EDMONDS  
 ; STREET: 1155 Avenue of the Americas  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: USA  
 ; ZIP: 10036-2711  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent in Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/646,695  
 ; FILING DATE: On Even Date Herewith  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Mistrock, S. Leslie  
 ; REGISTRATION NUMBER: 18,872  
 ; REFERENCE/DOCKET NUMBER: 6523-008  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (212) 790-9090  
 ; TELEFAX: (212) 869-9741/8864  
 ; TELEX: 66141 PENNIE  
 ; INFORMATION FOR SEQ ID NO: 6:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2109 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 US-08-646-695-6  
 Query Match 80.5%; Score 33; DB 3; Length 2109;  
 Best Local Similarity 71.4%; Pred. No. 2.8e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 HYAKNPI 7  
 DB 824 HFAENPI 830  
 RESULT 13  
 PCT-US96-06053-6  
 ; Sequence 6, Application PC/TUS9606053  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Yale University

; CURRENT APPLICATION NUMBER: US/09/107,532A  
 ; CURRENT FILING DATE: 2000-06-04  
 ; NUMBER OF SEQ ID NOS: 8252  
 ; SEQ ID NO 6604  
 ; LENGTH: 1627  
 ; TYPE: PRT  
 ; ORGANISM: Acinetobacter baumannii  
 US-09-328-352-6604  
 Query Match 80.5%; Score 33; DB 4; Length 1627;  
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 YAKNPI 7  
 DB 1487 YAKNPI 1492  
 RESULT 12  
 US-08-646-695-6  
 ; Sequence 6, Application US/08646695  
 ; Patent No. 6168943  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rose, John K.  
 ; TITLE OF INVENTION: RECOMBINANT VESICULOVIRUSES AND THEIR  
 ; TITLE OF INVENTION: USES  
 ; NUMBER OF SEQUENCES: 44  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: PENNIE & EDMONDS  
 ; STREET: 1155 Avenue of the Americas  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: USA  
 ; ZIP: 10036-2711  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent in Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/646,695  
 ; FILING DATE: On Even Date Herewith  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Mistrock, S. Leslie  
 ; REGISTRATION NUMBER: 18,872  
 ; REFERENCE/DOCKET NUMBER: 6523-008  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (212) 790-9090  
 ; TELEFAX: (212) 869-9741/8864  
 ; TELEX: 66141 PENNIE  
 ; INFORMATION FOR SEQ ID NO: 6:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2109 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 US-08-646-695-6  
 Query Match 80.5%; Score 33; DB 3; Length 2109;  
 Best Local Similarity 71.4%; Pred. No. 2.8e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 HYAKNPI 7  
 DB 824 HFAENPI 830  
 RESULT 13  
 PCT-US96-06053-6  
 ; Sequence 6, Application PC/TUS9606053  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Yale University

;; TITLE OF INVENTION: RECOMBINANT VESICULOVIRUSES AND THEIR  
;; TITLE OF INVENTION: USES  
;; NUMBER OF SEQUENCES: 41  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: PENNIE & EDMONDS  
;; STREET: 1155 Avenue of the Americas  
;; CITY: New York  
;; STATE: New York  
;; COUNTRY: USA  
;; ZIP: 10036-2711  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; OPERATING SYSTEM: IBM PC compatible  
;; SOFTWARE: Patent In Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: PCT/US96/06053  
;; FILING DATE: 01-MAY-1996  
;; CLASSIFICATION:  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Misrock, S. Leslie  
;; REGISTRATION NUMBER: 18,872  
;; REFERENCE/DOCKET NUMBER: 6523-009-228  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (212) 790-9090  
;; TELEFAX: (212) 869-9741/8864  
;; TELEX: 66141 PENNIE  
;; INFORMATION FOR SEQ ID NO: 6:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 2109 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
PCT-US96-06053-6

Query Match 80.5%; Score 33; DB 5; Length 2109;  
Best Local Similarity 71.4%; Pred. No. 2.8e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 HYAKNPI 7  
DB 824 HFAENPI 830

RESULT 14  
US-09-328-352-4692  
; Sequence 4692, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; FILE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 4692  
; LENGTH: 293  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-4692

Query Match 78.0%; Score 32; DB 4; Length 293;  
Best Local Similarity 83.3%; Pred. No. 52;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HYAKNP 6  
DB 159 HYKKNP 164

RESULT 15  
US-08-928-692-53

; Sequence 53, Application US/08928692  
; Patent No. 5958727  
; GENERAL INFORMATION:  
; APPLICANT: Brody, Howard  
; APPLICANT: Yaver, Deborah S.  
; APPLICANT: Lamea, Michael  
; APPLICANT: Hansen, Kim  
; TITLE OF INVENTION: Methods for Modifying the Production of  
; TITLE OF INVENTION: a Polypeptide  
; NUMBER OF SEQUENCES: 80  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 5958727 No. 5958727disk of No. 5958727th America, Inc.  
; STREET: 405 Lexington Avenue  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10174  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA: US/08/928,692  
; FILING DATE: 12-SEPT-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lambiris, Elias J  
; REGISTRATION NUMBER: 33,728  
; REFERENCE/DOCKET NUMBER: 4944.200-US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-867-0123  
; TELEFAX: 212-878-9655  
; INFORMATION FOR SEQ ID NO: 53:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 560 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: No. 5958727e  
US-08-928-692-53

Query Match 78.0%; Score 32; DB 2; Length 560;  
Best Local Similarity 57.1%; Pred. No. 1e+02;  
Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 HYAKNPI 7  
DB 101 HYSSNPV 107

Search completed: February 11, 2004, 17:13:33  
Job time : 11.4167 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 16:51:33 ; Search time 32.25 Seconds  
(without alignments)  
34.452 Million cell updates/sec

Title: US-09-901-187C-2

Perfect score: 41

Sequence: 1 HVXNFI 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_19Jun03.\*

- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*
- 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*
- 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*
- 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*
- 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*
- 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*
- 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*
- 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*
- 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*
- 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*
- 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*
- 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*
- 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*
- 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*
- 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*
- 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*
- 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*
- 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*
- 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*
- 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*
- 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*
- 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*
- 23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*
- 24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	100.0	7	AAE14547	Human alpha-synuclein
2	37	90.2	96	ABB55134	Lactococcus lactis
3	37	90.2	1669	ABB63980	Drosophila melanog
4	35	85.4	944	21 AAG29577	Arabidopsis thalia
5	35	85.4	973	21 AAG29576	Arabidopsis thalia
6	35	85.4	976	21 AAG40569	Arabidopsis thalia
7	35	85.4	989	21 AAG29575	Arabidopsis thalia
8	35	85.4	1005	21 AAG40568	Arabidopsis thalia
9	35	85.4	1021	21 AAG40567	Arabidopsis thalia

10	34	82.9	125	22	AAG78037	Piscirickettsia sa
11	34	82.9	136	22	AA001948	Human polypeptide
12	34	82.9	463	22	ABG29607	Novel human diagno
13	33	80.5	190	19	AAW98451	H. pylori GHPO 585
14	33	80.5	213	18	AAW55457	H. pylori ORF 029P
15	33	80.5	214	20	AAV17186	H. pylori outer me
16	33	80.5	230	18	AAW55265	H. pylori ORF 06ge
17	33	80.5	296	22	ABG04955	Novel human diagno
18	33	80.5	524	22	ABG59240	SHP-2 protein. Un
19	33	80.5	593	15	AAW52991	Human protein-tyro
20	33	80.5	593	17	AAW99313	Human SH-PTP2 (pro
21	33	80.5	593	20	AAV13476	Peptide seq ID No:
22	33	80.5	593	22	AAW59213	SHP-2 mutant E76A
23	33	80.5	593	22	AAW59214	SHP-2 protein. Un
24	33	80.5	593	22	AAW59215	SHP-2 activated do
25	33	80.5	593	22	AAW59218	SHP-2 activated do
26	33	80.5	593	22	AAW59219	SHP-2 activated do
27	33	80.5	593	22	AAW59220	SHP-2 activated do
28	33	80.5	593	22	AAW59221	SHP-2 activated tr
29	33	80.5	593	22	AAW59222	SHP-2 activated tr
30	33	80.5	593	22	AAW59223	SHP-2 activated qu
31	33	80.5	593	22	AAW59224	SHP-2 activated qu
32	33	80.5	593	22	AAW59225	SHP-2 activated qu
33	33	80.5	593	22	AAW59226	SHP-2 activated qu
34	33	80.5	593	22	AAW59227	SHP-2 activated qu
35	33	80.5	597	20	AAV13475	Peptide Seq ID No:
36	33	80.5	597	23	ABG59480	Human bait protein
37	33	80.5	611	22	ABG04956	Novel human diagno
38	33	80.5	790	24	ABP76733	Streptomyces virid
39	33	80.5	1753	22	AAW20291	Vesicular stomatit
40	33	80.5	2109	17	AAW04540	Vesiculovirus larg
41	33	80.5	2109	22	AAW20292	Vesicular stomatit
42	33	80.5	2109	22	AAW59299	Vesicular stomatit
43	33	80.5	19938	24	ABB98398	Streptomyces virid
44	32	78.0	31	22	ABB95982	Human testicular a
45	32	78.0	31	22	AAW95285	Human reproductive

ALIGNMENTS

RESULT 1

AAE14547  
ID AAE14547 standard; peptide; 7 AA.

AC AAE14547;

XX 17-MAY-2002 (first entry)

DE Human alpha-synuclein aggregation inhibitor #2.

XX Alpha-synuclein; inhibitor; neurodegenerative disease; Lewy body;  
KW Parkinson's disease; Alzheimer's disease; diffuse Lewy body disease;  
KW multiple system atrophy; Hallervorden-Spatz disease; human.

XX Homo-sapiens

XX WO200204482-A1

XX 17-JAN-2002.

XX 06-JUL-2001; 2001WO-US21379.

XX 07-JUL-2000; 2000US-217319P.

XX 28-MAR-2001; 2001US-279199P.

XX (PANA-) PANACEA PHARM INC.

XX Wlozoin B, Ostretova-Golts N, Lebowitz MS;

XX WPI; 2002-179695/23.

XX Determination of an agent capable of inhibiting aggregation of alpha



PT synuclein useful for treating a neurodegenerative disease involves  
PT determining aggregation of alpha synuclein in the presence of exogenous  
PT iron or copper  
XX  
XX Claim 40; Page 37; 52pp; English.  
XX  
XX The invention relates to screening of inhibitors of alpha-synuclein  
CC aggregation in the presence of exogenous iron or copper. The inhibitors  
CC are magnesium and alpha-synuclein binding peptides, which are  
CC useful for treating neurodegenerative disease that involves  
CC the formation of Lewy bodies e.g. Parkinson's disease (PD), Alzheimer's  
CC disease (AD), diffuse Lewy body disease, mixed AD-PD, multiple system  
CC atrophy and Hallervorden-Spatz disease. The present sequence is a  
CC peptide that binds to C-terminal portion of human alpha-synuclein and  
CC inhibits its aggregation.  
XX  
SQ Sequence 7 AA;  
  
Query Match 100.0%; Score 41; DB 23; Length 7;  
Best Local Similarity 100.0%; Pred. No. 9.3e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 HYAKNPI 7  
D5 |||||  
D5 1 HYAKNPI 7  
  
RESULT 2  
ABB55134  
ID ABB55134 standard; Protein; 96 AA.  
XX  
AC ABB55134;  
XX  
DT 16-MAY-2002 (first entry)  
XX  
DE Lactococcus lactis protein yseD.  
XX  
KW Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.  
XX  
OS Lactococcus lactis IL1403.  
XX  
PN FR2807446-A1.  
XX  
PD 12-OCT-2001.  
XX  
PF 11-APR-2000; 2000FR-0004630.  
XX  
PR 11-APR-2000; 2000FR-0004630.  
XX  
PA (INRG ) INRA INST NAT RECH AGRONOMIQUE.  
XX  
PI Bolotine A, Sorokine A, Renault P, Ehrlich SD;  
XX  
DR WPI; 2002-043418/06.  
XX  
PT New nucleotide sequence useful in the identification or Lactococcus  
PT lactic and related species -  
XX  
XX Claim 6; SEQ ID No 1836; 2504pp; French.  
XX  
XX The present invention is related to a Lactococcus lactis nucleotide  
CC sequence (ABA90521) and related proteins (ABB53300-ABB55621). The  
CC nucleic acid sequence is useful in the detection and/or amplification of  
CC nucleic acid sequence, particularly to identify Lactococcus lactis or  
CC related species. The proteins of the invention are useful for the  
CC biosynthesis or biodegradation of a composition of interest. The  
CC invention helps research in lactic bacteria, particularly useful in the  
CC production of yogurt and cheese.  
CC Note: The sequence data for this patent is based on equivalent patent  
CC WO200177334 (published 18-OCT-2001) which is available in electronic  
CC format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 96 AA;

Query Match 90.2%; Score 37; DB 23; Length 96;  
Best Local Similarity 85.7%; Pred. No. 5.8;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 HYAKNPI 7  
D5 |||||  
D5 26 HYAKNPI 32  
  
RESULT 3  
ABB63980  
ID ABB63980 standard; Protein; 1669 AA.  
XX  
AC ABB63980;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 18732.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEXE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR WPI; 2001-856860/75.  
XX  
DR N-P8DB; ABL08083.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
XX Disclosure; SEQ ID NO 18732; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABLO1840-ABLI16175) and the encoded proteins  
CC (ABBS7737-ABBY2072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 1669 AA;  
  
Query Match 90.2%; Score 37; DB 22; Length 1669;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 HYAKNP 6  
D5 |||||  
D5 155 HYAKNP 160  
  
RESULT 4  
AAG29577  
ID AAG29577 standard; Protein; 944 AA.

XX AAG29577;  
AC 17-OCT-2000 (first entry)  
XX  
DT  
XX  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 35215.  
DE  
XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; Genetic mapping; Gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
XX EP1033405-A2.  
PN  
XX  
XX 06-SEP-2000.  
PD  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
PF  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130691.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 18-MAY-1999; 99US-0134370.  
PR 19-MAY-1999; 99US-0134768.  
PR 20-MAY-1999; 99US-0134941.  
PR 21-MAY-1999; 99US-0135124.  
PR 24-MAY-1999; 99US-0135353.  
PR 25-MAY-1999; 99US-0135629.  
PR 27-MAY-1999; 99US-0136021.  
PR 01-JUN-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 03-JUN-1999; 99US-0137222.  
PR 04-JUN-1999; 99US-0137528.  
PR 07-JUN-1999; 99US-0137502.  
PR 08-JUN-1999; 99US-0137724.  
PR 10-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 14-JUN-1999; 99US-0138847.  
PR 16-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 18-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139464.  
PR 18-JUN-1999; 99US-0139465.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 04-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 06-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148585.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149829.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.

PR 25-AUG-1999; 99US-0150566.  
 PR 26-AUG-1999; 99US-0150884.  
 PR 27-AUG-1999; 99US-0151065.  
 PR 27-AUG-1999; 99US-0151066.  
 PR 27-AUG-1999; 99US-0151080.  
 PR 30-AUG-1999; 99US-0151303.  
 PR 31-AUG-1999; 99US-0151438.  
 PR 01-SEP-1999; 99US-0151930.  
 PR 07-SEP-1999; 99US-0152363.  
 PR 10-SEP-1999; 99US-0153070.  
 PR 13-SEP-1999; 99US-0153758.  
 PR 15-SEP-1999; 99US-0154018.  
 PR 16-SEP-1999; 99US-0154039.  
 PR 20-SEP-1999; 99US-0154779.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161820.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 85.4%; Score 35; DB 21; Length 944;  
 Best Local Similarity 85.7%; Pred. No. 1.9e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HYAKNPI 7  
 DB 199 HYAHNPI 205

RESULT 5  
 AAG29576  
 ID AAG29576 standard; Protein; 973 AA.  
 XX  
 AC AAG29576;  
 XX  
 DT 17-OCT-2000 (first entry)  
 XX  
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 35214.  
 XX

KW Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 XX termination sequence.  
 OS Arabidopsis thaliana.  
 XX EP1033405-A2.  
 XX  
 PD 06-SEP-2000.  
 XX  
 PF 25-FEB-2000; 2000EP-0301439.  
 XX  
 PR 25-FEB-1999; 99US-0121825.  
 PR 05-MAR-1999; 99US-0121825.  
 PR 09-MAR-1999; 99US-0123548.  
 PR 23-MAR-1999; 99US-0125788.  
 PR 25-MAR-1999; 99US-0126264.  
 PR 29-MAR-1999; 99US-0126785.  
 PR 01-APR-1999; 99US-0127462.  
 PR 06-APR-1999; 99US-0128234.  
 PR 08-APR-1999; 99US-0128714.  
 PR 16-APR-1999; 99US-0129845.  
 PR 19-APR-1999; 99US-0130077.  
 PR 21-APR-1999; 99US-0130449.  
 PR 23-APR-1999; 99US-0130510.  
 PR 23-APR-1999; 99US-0130891.  
 PR 28-APR-1999; 99US-0131449.  
 PR 30-APR-1999; 99US-0132048.  
 PR 30-APR-1999; 99US-0132407.  
 PR 04-MAY-1999; 99US-0132484.  
 PR 05-MAY-1999; 99US-0132485.  
 PR 06-MAY-1999; 99US-0132486.  
 PR 06-MAY-1999; 99US-0132487.  
 PR 07-MAY-1999; 99US-0132863.  
 PR 11-MAY-1999; 99US-0134256.  
 PR 14-MAY-1999; 99US-0134218.  
 PR 14-MAY-1999; 99US-0134219.  
 PR 14-MAY-1999; 99US-0134219.  
 PR 14-MAY-1999; 99US-0134221.  
 PR 14-MAY-1999; 99US-0134370.  
 PR 18-MAY-1999; 99US-0134768.  
 PR 19-MAY-1999; 99US-0134941.  
 PR 20-MAY-1999; 99US-0135124.  
 PR 21-MAY-1999; 99US-0135353.  
 PR 24-MAY-1999; 99US-0135629.  
 PR 25-MAY-1999; 99US-0136021.  
 PR 27-MAY-1999; 99US-0136392.  
 PR 28-MAY-1999; 99US-0136782.  
 PR 01-JUN-1999; 99US-0137222.  
 PR 03-JUN-1999; 99US-0137528.  
 PR 04-JUN-1999; 99US-0137502.  
 PR 07-JUN-1999; 99US-0137724.  
 PR 08-JUN-1999; 99US-0138094.  
 PR 10-JUN-1999; 99US-0138540.  
 PR 10-JUN-1999; 99US-0138847.  
 PR 14-JUN-1999; 99US-0139119.  
 PR 16-JUN-1999; 99US-0139452.  
 PR 16-JUN-1999; 99US-0139453.  
 PR 17-JUN-1999; 99US-0139492.  
 PR 18-JUN-1999; 99US-0139454.  
 PR 18-JUN-1999; 99US-0139455.  
 PR 18-JUN-1999; 99US-0139456.  
 PR 18-JUN-1999; 99US-0139457.  
 PR 18-JUN-1999; 99US-0139458.  
 PR 18-JUN-1999; 99US-0139459.  
 PR 18-JUN-1999; 99US-0139460.  
 PR 18-JUN-1999; 99US-0139461.  
 PR 18-JUN-1999; 99US-0139462.  
 PR 18-JUN-1999; 99US-0139463.  
 PR 18-JUN-1999; 99US-0139750.  
 PR 18-JUN-1999; 99US-0139763.  
 PR 21-JUN-1999; 99US-0139817.  
 PR 22-JUN-1999; 99US-0139859.  
 PR 23-JUN-1999; 99US-0140353.

PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.

PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158028.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 85.4%; Score 35; DB 21; Length 973;  
Best Local Similarity 85.7%; Pred. No. 1.9e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HYAKUPI 7  
Db 228 HYAHNPI 234

RESULT 6  
AAG40569  
ID AAG40569 standard; Protein; 976 AA.  
XX  
AC AAG40569;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 50353.

XX Protein identification; signal transduction pathway; metabolic pathway;  
XX hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.  
XX Arabidopsis thaliana.  
XX  
PN EP1033405-A2.

XX	06-SEP-2000.	PR	02-JUL-1999;	99US-0142055.
PD		PR	06-JUL-1999;	99US-0142390.
XX		PR	08-JUL-1999;	99US-0142803.
XX		PR	09-JUL-1999;	99US-0142920.
PF	25-FEB-2000; 2000EP-0301439.	PR	12-JUL-1999;	99US-0142977.
XX		PR	13-JUL-1999;	99US-0143542.
PR	25-FEB-1999;	PR	14-JUL-1999;	99US-0143624.
PR	05-MAR-1999;	PR	15-JUL-1999;	99US-0144005.
PR	09-MAR-1999;	PR	16-JUL-1999;	99US-0144085.
PR	23-MAR-1999;	PR	16-JUL-1999;	99US-0144086.
PR	25-MAR-1999;	PR	19-JUL-1999;	99US-0144325.
PR	29-MAR-1999;	PR	19-JUL-1999;	99US-0144331.
PR	01-APR-1999;	PR	19-JUL-1999;	99US-0144332.
PR	06-APR-1999;	PR	19-JUL-1999;	99US-0144333.
PR	08-APR-1999;	PR	19-JUL-1999;	99US-0144334.
PR	16-APR-1999;	PR	19-JUL-1999;	99US-0144335.
PR	19-APR-1999;	PR	20-JUL-1999;	99US-0144352.
PR	21-APR-1999;	PR	20-JUL-1999;	99US-0144632.
PR	23-APR-1999;	PR	20-JUL-1999;	99US-0144632.
PR	23-APR-1999;	PR	20-JUL-1999;	99US-0144884.
PR	28-APR-1999;	PR	21-JUL-1999;	99US-0144884.
PR	30-APR-1999;	PR	21-JUL-1999;	99US-0145086.
PR	30-APR-1999;	PR	21-JUL-1999;	99US-0145086.
PR	30-APR-1999;	PR	22-JUL-1999;	99US-0145088.
PR	04-MAY-1999;	PR	22-JUL-1999;	99US-0145085.
PR	05-MAY-1999;	PR	22-JUL-1999;	99US-0145087.
PR	06-MAY-1999;	PR	22-JUL-1999;	99US-0145089.
PR	06-MAY-1999;	PR	22-JUL-1999;	99US-0145192.
PR	07-MAY-1999;	PR	23-JUL-1999;	99US-0145145.
PR	11-MAY-1999;	PR	23-JUL-1999;	99US-0145216.
PR	14-MAY-1999;	PR	23-JUL-1999;	99US-0145224.
PR	14-MAY-1999;	PR	26-JUL-1999;	99US-0145276.
PR	14-MAY-1999;	PR	27-JUL-1999;	99US-0145913.
PR	18-MAY-1999;	PR	27-JUL-1999;	99US-0145918.
PR	19-MAY-1999;	PR	27-JUL-1999;	99US-0145919.
PR	20-MAY-1999;	PR	28-JUL-1999;	99US-0145951.
PR	21-MAY-1999;	PR	28-JUL-1999;	99US-0146386.
PR	24-MAY-1999;	PR	02-AUG-1999;	99US-0146388.
PR	25-MAY-1999;	PR	02-AUG-1999;	99US-0146389.
PR	27-MAY-1999;	PR	03-AUG-1999;	99US-0147038.
PR	28-MAY-1999;	PR	04-AUG-1999;	99US-0147204.
PR	01-JUN-1999;	PR	05-AUG-1999;	99US-0147192.
PR	03-JUN-1999;	PR	05-AUG-1999;	99US-0147260.
PR	04-JUN-1999;	PR	06-AUG-1999;	99US-0147303.
PR	07-JUN-1999;	PR	06-AUG-1999;	99US-0147416.
PR	08-JUN-1999;	PR	09-AUG-1999;	99US-0147493.
PR	10-JUN-1999;	PR	09-AUG-1999;	99US-0147935.
PR	10-JUN-1999;	PR	10-AUG-1999;	99US-0148171.
PR	14-JUN-1999;	PR	11-AUG-1999;	99US-0148319.
PR	16-JUN-1999;	PR	12-AUG-1999;	99US-0148341.
PR	16-JUN-1999;	PR	13-AUG-1999;	99US-0148565.
PR	17-JUN-1999;	PR	13-AUG-1999;	99US-0148684.
PR	18-JUN-1999;	PR	16-AUG-1999;	99US-0149368.
PR	18-JUN-1999;	PR	17-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	PR	18-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	PR	20-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	PR	20-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	PR	20-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	PR	23-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	PR	25-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	PR	26-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	PR	27-AUG-1999;	99US-0151065.
PR	18-JUN-1999;	PR	27-AUG-1999;	99US-0151066.
PR	18-JUN-1999;	PR	27-AUG-1999;	99US-0151080.
PR	18-JUN-1999;	PR	30-AUG-1999;	99US-0151303.
PR	21-JUN-1999;	PR	31-AUG-1999;	99US-0151438.
PR	22-JUN-1999;	PR	01-SEP-1999;	99US-0151930.
PR	23-JUN-1999;	PR	07-SEP-1999;	99US-0152363.
PR	24-JUN-1999;	PR	10-SEP-1999;	99US-0153070.
PR	24-JUN-1999;	PR	13-SEP-1999;	99US-0153758.
PR	28-JUN-1999;	PR	15-SEP-1999;	99US-0154018.
PR	30-JUN-1999;	PR	16-SEP-1999;	99US-0154039.
PR	01-JUL-1999;	PR	20-SEP-1999;	99US-0154779.

```
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 85.4%; Score 35; DB 21; Length 976;
Best Local Similarity 85.7%; Pred. No. 1.9e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HYAKNPI 7
DB 211 HYAHNPI 217

RESULT 7
ID AAG29575 standard; Protein; 989 AA.
XX AC AAG29575;
XX DT 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 35213.
XX DE Arabidopsis thaliana.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
XX PR 05-MAR-1999; 99US-0123180.
```

```
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144332.
PR 20-JUL-1999; 99US-0144332.
PR 20-JUL-1999; 99US-0144332.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145224.
PR 27-JUL-1999; 99US-0145213.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147202.
PR 05-AUG-1999; 99US-0147192.
PR 08-AUG-1999; 99US-0147260.
PR 08-AUG-1999; 99US-0147303.
PR 08-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 07-SEP-1999; 99US-0151930.
PR 10-SEP-1999; 99US-0152363.
PR 13-SEP-1999; 99US-0153070.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.

PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159283.
PR 13-OCT-1999; 99US-0159284.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0158584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 85.4%; Score 35; DB 21; Length 989;
Best Local Similarity 85.7%; Pred. No. 2e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HYAKNPI 7
Db 244 HYAHNPI 250

RESULT 8
AAG40568
ID AAG40568 standard; Protein; 1005 AA.
XX
AC AAG40568;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 50352.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0123788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
```

PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 14-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 14-JUN-1999; 99US-0138847.  
PR 16-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 21-JUN-1999; 99US-0139763.  
PR 22-JUN-1999; 99US-0139817.  
PR 23-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 24-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 29-JUN-1999; 99US-0140823.  
PR 30-JUN-1999; 99US-0140991.  
PR 01-JUL-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 06-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-APR-1999; 99US-0129845.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145132.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 26-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147132.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155133.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158212.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.



PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159337.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 21-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 85.4%; Score 35; DB 21; Length 1005;

Best Local Similarity 85.7%; Pred. No. 2e+02; Indels 0; Gaps 0;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 HYAKNPI 7  
||| |||  
Db 240 HYAHNPI 246

RESULT 9  
AAG40567  
ID AAG40567 standard; Protein; 1021 AA.

XX AC AAG40567;  
XX DT 18-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 50351.

XX KW Protein identification; signal transduction pathway; metabolic pathway;  
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX KW termination sequence.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 99US-0121825.

XX PR 05-MAR-1999; 99US-0123180.

XX PR 09-MAR-1999; 99US-0123548.

XX PR 23-MAR-1999; 99US-0125788.

XX PR 25-MAR-1999; 99US-0126264.

XX PR 29-MAR-1999; 99US-0126785.

XX PR 01-APR-1999; 99US-0127462.

XX PR 06-APR-1999; 99US-0128234.

XX PR 18-APR-1999; 99US-0128714.

XX PR 16-APR-1999; 99US-0128845.

XX PR 19-APR-1999; 99US-0130077.

XX PR 21-APR-1999; 99US-0130449.

XX PR 23-APR-1999; 99US-0130510.

XX PR 28-APR-1999; 99US-0130891.

XX PR 30-APR-1999; 99US-0131449.

XX PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 18-JUL-1999; 99US-0144086.  
PR 18-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.

```
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 09-AUG-1999; 99US-0149175.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149923.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152383.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158039.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.

PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160880.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 85.4%; Score 35; DB 21; Length 1021;
Best Local Similarity 85.7%; Pred. NO. 2e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HYAKNPI 7
Db 256 HYAHNPI 262

RESULT 10
AAG78037
ID AAG78037 standard; Protein; 125 AA.
XX AC AAG78037;
XX AC AAG78037;
XX DT 15-JAN-2002 (first entry)
XX DE Piscirickettsia salmonis polynucleotide clone 15.
XX KW Piscirickettsia salmonis; piscirickettsiosis; salmonid rickettsial;
XX KW septicaemia; SRS; surface antigen; vaccine; antibacterial; fish; ss.
XX OS Piscirickettsia salmonis.
XX FH Key Location/Qualifiers
FT Misc-difference 3 /label= unknown
FT /note= "Encoded by NGA"
XX WO200168865-A2.
XX PD 20-SEP-2001.
XX PF 12-MAR-2001; 2001WO-GB01055.
XX PR 11-MAR-2000; 2000GB-0005838.
XX PR 01-JUL-2000; 2000GB-0016080.
XX PR 01-JUL-2000; 2000GB-0016082.
XX PR 29-JUL-2000; 2000GB-0018599.
XX PA (AQUA-) AQUA HEALTH EURO LTD.
XX PI Simard N, Brouwers H, Jones S, Griffiths S, Valenzuela P;
XX PI Burzio L;
XX WPI; 2001-639050/73.
XX DR N-P8DB; AAI68358.
XX PT New nucleic acids encoding an amino acid sequence homologous to the
XX PT surface antigen present on piscirickettsia salmonis are useful to
XX PT protect fish against piscirickettsiosis -
XX PS Claim 6; Fig 17; 25pp; English.
XX
```

CC The invention relates to nucleic acid sequences and the encoded protein  
 CC of a least part of the surface antigen present on Piscirickettsia  
 CC salmonis for production of a vaccine with antibacterial activity to  
 CC protect fish against P. salmonis which causes piscirickettsiosis, also  
 CC known as salmonid rickettsial septicaemia.

XX  
 SQ Sequence 125 AA;

Query Match 82.9%; Score 34; DB 22; Length 125;  
 Best Local Similarity 71.4%; Pred. No. 32;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 HYAKNPI 7  
 :|||||:  
 Db 84 YYAKNPV 90

RESULT 11  
 AAO01948  
 ID AAO01948 standard; Protein; 136 AA.  
 XX  
 AC AAO01948;  
 XX  
 DT 06-NOV-2001 (first entry)  
 XX  
 DE Human polypeptide SEQ ID NO 15840.  
 XX  
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorders; arthritis; inflammation.  
 XX  
 OS Homo sapiens.  
 XX  
 WO200164835-A2.  
 PN  
 XX  
 PD 07-SEP-2001.  
 XX  
 PF 26-FEB-2001; 2001WO-US04927.  
 XX  
 PR 28-FEB-2000; 2000US-0515126.  
 PR 18-MAY-2000; 2000US-0577409.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Liu C, Drmanac RT;  
 XX  
 WPI; 2001-514838/56.  
 DR N-PSDB; AA181879.  
 XX  
 PT Isolated nucleic acids and polypeptides, useful for preventing  
 PT diagnosing and treating e.g. leukaemia, inflammation and immune  
 PT disorders -  
 XX  
 PS Claim 20; SEQ ID NO 15840; 1399pp + Sequence Listing; English.  
 XX  
 CC The invention relates to human polynucleotides (AA179941-AA193841) and  
 CC the encoded proteins (AA000010-AA013910) that exhibit activity relating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulatory  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX  
 SQ Sequence 136 AA;

Query Match 82.9%; Score 34; DB 22; Length 136;  
 Best Local Similarity 83.3%; Pred. No. 35;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 HYAKNP 6  
 :|||||:  
 Db 3 HYSKNP 8

RESULT 12  
 ABG29607  
 ID ABG29607 standard; Protein; 463 AA.  
 XX  
 AC ABG29607;  
 XX  
 DT 18-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #29598.  
 XX  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 WO200175067-A2.  
 PN  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YT;  
 XX  
 WPI; 2001-639362/73.  
 DR N-PSDB; AAS93794.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX  
 PS Claim 20; SEQ ID No 59966; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG3037 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX  
 SQ Sequence 463 AA;

Query Match 82.9%; Score 34; DB 22; Length 463;  
 Best Local Similarity 71.4%; Pred. No. 1.4e+02;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 HYAKNPI 7  
148 HNASNPL 154

RESULT 13  
AAW98451  
ID AAW98451 standard; Protein; 190 AA.  
XX AC AAW98451;  
XX 31-MAR-1999 (first entry)  
DT H. pylori GHPO 585 protein.  
DE GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;  
XX peptic ulcer disease.  
KW Helicobacter pylori.  
XX OS W09843478-A1.  
XX PN 08-OCT-1998.  
XX PD 01-APR-1998; 98WO-US06371.  
XX PF 29-JUL-1997; 97US-0902615.  
XX PR 01-APR-1997; 97US-0833457.  
XX PR 24-JUN-1997; 97US-0881227.  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.  
XX PI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;  
XX WPI; 1998-542293/46.  
XX DR N-PSDB; AAX14170.  
XX PT New isolated Helicobacter polynucleotides - used to develop products  
XX for the diagnosis, prevention and treatment of Helicobacter  
XX infections and gastrointestinal diseases  
XX Claim 8; Page 787-788; 2054pp; English.

This sequence represents a Helicobacter pylori GHPO protein of the  
invention. The polypeptides can be used for preventing or treating  
Helicobacter infections, and gastroduodenal diseases associated with  
these infections, including acute, chronic, and atrophic gastritis, and  
peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be  
used for the production of antibodies. The products can also be used for  
detection and diagnosis.

Sequence 190 AA;  
Query Match 80.5%; Score 33; DB 19; Length 190;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 YAKNPI 7  
91 YAKNPI 96

RESULT 14  
AAW55457  
ID AAW55457 standard; Protein; 213 AA.  
XX AC AAW55457;  
XX 24-JUN-1998 (first entry)  
DT XX

H. pylori ORF 02gp20706\_23639775\_f3\_32 cell envelope OMP.

Cytoplasmic; vaccine; prevention; treatment; infection; envelope;  
identification; binding compound; bacteria; life cycle; activator;  
inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.

Helicobacter pylori.

W09737044-A1.

09-OCT-1997.

27-MAR-1997; 97WO-US05223.

06-DEC-1996; 96US-0761318.

29-MAR-1996; 96US-0625811.

02-APR-1996; 96US-0758731.

25-OCT-1996; 96US-0736905.

28-OCT-1996; 96US-0738859.

(ASTR) ASTRA AB.

Alm RA, Smith D;

WPI; 1997-503122/46.

N-PSDB; AAV24866.

Helicobacter pylori nucleic acid sequences and encoded

polypeptide(s) - useful in vaccines to treat or prevent H. pylori

infection and for diagnosis of H. pylori infection

Claims 14,80; Page 664; 1145pp; English.

This sequence is a H. pylori cell envelope outer membrane protein

(OMP) having a terminal Phe residue.

The protein may be used in a vaccine to prevent or treat H. pylori

infection or to identify H. pylori polypeptide binding compounds,

useful as potential H. pylori life cycle activators or inhibitors. The

DNA and probes derived from it may be used for the identification of

H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic

acid sequences complementary to the DNA act as antisense sequences and

can be used to prevent the translation of H. pylori mRNA. Antibodies

against the protein can be used in immunoassays to evaluate the abundance

and distribution of H. pylori-specific antigens. The genomic sequence of

H. pylori (ATCC 55679) was determined from overlapping contigs generated

by mechanically shearing the bacterial DNA. The sequences were analysed

for ORF of at least 180 nucleotides, and the predicted coding regions

defined by computer evaluation. To identify likely H. pylori antigens for

vaccine development, the amino acid sequences predicted from various ORF

were analysed for significant homology to other known or exported

membrane proteins. Having identified and determined the sequences of

interest, particular regions can be isolated from H. pylori by PCR

amplification for recombinant polypeptide production, e.g. in E. coli

hosts.

Sequence 213 AA;

Query Match 80.5%; Score 33; DB 18; Length 213;

Best Local Similarity 100.0%; Pred. No. 93;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 YAKNPI 7

114 YAKNPI 119

RESULT 15  
AAV17186  
ID AAV17186 standard; Protein; 214 AA.  
XX AC AAV17186;  
XX 03-AUG-1999 (first entry)  
DT XX

XX H. pylori outer membrane polypeptide.  
DE  
XX Outer membrane polypeptide; OMP; vaccine; H. pylori infection; humoral;  
KW cellular immune response.  
XX  
XX Helicobacter pylori.  
XX  
XX WO9921959-A2.  
XX  
XX 06-MAY-1999.  
XX  
XX 28-OCT-1998; 98WO-US22883.  
XX  
XX 17-DEC-1997; 97US-0993001.  
PR 28-OCT-1997; 97US-0959131.  
XX  
XX (GENO-) GENOME THERAPEUTICS CORP.  
PA  
XX Alm RA, Ellis RW, Guild BC, Noonan BM, Smith D;  
PI WPI; 1999-326698/27.  
XX N-PSDB; AAX75805.  
DR  
XX Cellular vaccine against Helicobacter pylori  
PT  
XX Claim 7; Page 252-253; 352pp; English.  
PS  
XX The invention relates to a vaccine for preventing or treating infections  
CC by Helicobacter pylori. The vaccine contains at least one isolated  
CC H. pylori polypeptide, or its fragments, in a carrier, where the  
CC carrier is a Salmonella, Vibrio cholerae or Shigella vector containing a  
CC nucleic acid encoding the H. pylori polypeptide. The vaccines induce  
CC humoral and cellular immune responses. The vaccines are used to treat or  
CC prevent infections by H. pylori. Sequences AAX75779 to AAX75837 represent  
CC nucleic acid sequences encoding H. pylori outer membrane polypeptides  
CC (OMPs) AAY17160 to AAY17218.  
XX  
XX Sequence 214 AA;  
SQ  
Query Match 80.5%; Score 33; DB 20; Length 214;  
Best Local Similarity 100.0%; Pred. NO. 93;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 2 YAKNPI 7  
DB 115 YAKNPI 120  
|||||

Search completed: February 11, 2004, 17:02:45  
Job time : 33.25 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model  
Run on: February 11, 2004, 16:52:34 ; Search time 24.5833 Seconds  
(without alignments)  
73.479 Million cell updates/sec

Title: US-09-901-187C-3  
Perfect score: 32  
Sequence: 1 ATINKSL 7  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5  
Searched: 830525 seqs, 258052604 residues  
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTRMBL\_23:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mmc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriaph:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	32	100.0	118	5 Q00084	Q00084 drosophila
2	30	93.8	272	10 Q9FK73	Q9FK73 arabidopsis
3	29	90.6	312	16 Q92NH6	Q92NH6 rhizobium m
4	29	90.6	433	16 Q8R3L4	Q8R3L4 thermosacchar
5	29	90.6	1516	3 Q94477	Q94477 schizosacch
6	28	87.5	126	2 Q8KQV8	Q8KQV8 vibrio chol
7	28	87.5	127	3 Q8N1L4	Q8N1L4 schizosacch
8	28	87.5	195	6 Q9XS26	Q9XS26 bos taurus
9	28	87.5	245	6 Q9GK14	Q9GK14 bos taurus
10	28	87.5	289	5 Q9N4M9	Q9N4M9 caenorhabdi
11	28	87.5	368	12 Q69065	Q69065 human herpe
12	28	87.5	368	12 Q69463	Q69463 human herpe
13	28	87.5	387	16 Q8DW81	Q8DW81 streptococc
14	28	87.5	388	12 Q9PV12	Q9PV12 human cytom
15	28	87.5	396	2 Q30859	Q30859 rhodobacter
16	28	87.5	403	12 Q69052	Q69052 human herpe

17	28	87.5	439	10 Q9LY67	Q9LY67 arabidopsis
18	28	87.5	485	10 Q8GW44	Q8GW44 arabidopsis
19	28	87.5	469	10 Q8LEG9	Q8LEG9 arabidopsis
20	28	87.5	550	5 Q44180	Q44180 caenorhabdi
21	28	87.5	552	6 Q95M29	Q95M29 oryctolagus
22	28	87.5	552	13 Q9DG41	Q9DG41 brachydanio
23	28	87.5	553	4 Q96RV2	Q96RV2 homo sapien
24	28	87.5	554	11 Q8CH92	Q8CH92 rattus norv
25	28	87.5	554	11 Q8CEB9	Q8CEB9 mus musculu
26	28	87.5	605	16 Q8DJ58	Q8DJ58 synchococc
27	28	87.5	636	16 Q9RZHS	Q9RZHS deinococcus
28	28	87.5	642	8 Q8SLR5	Q8SLR5 clethra bar
29	28	87.5	814	5 Q9VSW5	Q9VSW5 drosophila
30	28	87.5	814	5 P91945	P91945 drosophila
31	28	87.5	1203	5 Q8WPL4	Q8WPL4 cikopleura
32	28	87.5	1337	5 Q8IC23	Q8IC23 plasmodium
33	28	87.5	1556	3 Q06554	Q06554 saccharomyc
34	28	87.5	1588	3 Q9UT79	Q9UT79 schizosacch
35	28	87.5	2371	5 Q8SS51	Q8SS51 encephalito
36	28	87.5	2756	10 Q9LJ60	Q9LJ60 arabidopsis
37	28	87.5	2833	2 Q8VPL1	Q8VPL1 pasteurella
38	27	84.4	112	2 Q93DV7	Q93DV7 unidentified
39	27	84.4	134	1 Q9C4N2	Q9C4N2 methanosarc
40	27	84.4	160	16 Q9CK10	Q9CK10 pasteurella
41	27	84.4	225	16 Q8VYN7	Q8VYN7 anabaena sp
42	27	84.4	228	5 Q8ID89	Q8ID89 plasmodium
43	27	84.4	256	5 Q8I6M1	Q8I6M1 eptaretus
44	27	84.4	257	5 Q8I6M4	Q8I6M4 ephydatia f
45	27	84.4	258	5 Q8I6M2	Q8I6M2 eptaretus

## ALIGNMENTS

RESULT 1  
Q00084  
ID Q00084 PRELIMINARY; PRT; 118 AA.  
AC Q00084;  
DT 01-NOV-1996 (TRENBLrel. 01, Created)  
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
DE Kinesin-like protein (KLP67A) (Fragment).  
GN KLP67A OR CGI0923.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=DP CN BW;  
RX MEDLINE=92020874; PubMed=1924306;  
RA Stewart R.J., Pesavento P.A., Woarpe D.N., Goldstein L.S.;  
RT "Identification and partial characterization of six members of the  
kinesin superfamily in Drosophila."  
RL Proc. Natl. Acad. Sci. U.S.A. 88:8470-8474(1991).  
CC -!- FUNCTION: POSSIBLE MICROTUBULE-DEPENDENT MECHANOCHEMICAL ENZYME.  
CC -!- SIMILARITY: BELONGS TO A FAMILY OF KINESIN-LIKE PROTEINS  
CHARACTERIZED BY THE PRESENCE OF MECHANOCHEMICAL DOMAIN TETHERED  
TO DIFFERENT PROTEIN BINDING DOMAINS.  
CC EMBL; W74429; AA28656.1; -;  
DR HSP; P56536; 2KIN.  
DR Flybase; FBgn004379; Klp67A.  
DR InterPro; IPR001752; kinesin\_motor.  
DR Pfam; PF00225; kinesin; 1.  
DR PRINTS; PR00380; KINESINHEAVY.  
DR SMART; SM00129; KISC1.  
DR PROSITE; PS50067; KINESIN MOTOR DOMAIN2; 1.  
KW ATP-binding; Coiled coil; Microtubules; Motor protein.  
FT NON TER 1 1  
FT NON TER 118 118  
SQ SEQUENCE 118 AA; 13076 MW; D2EA085C1267B2E9 CRC64;

Query Match 100.0%; Score 32; DB 5; Length 118;  
Best Local Similarity 100.0%; Pred. No. 9.3;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATINKSL 7  
Db 52 ATINKSL 58

# RESULT 2

Q9FK73 Q9FK73 PRELIMINARY; PRT; 272 AA.  
AC Q9FK73; (TREMELrel. 15, Created)  
DT 01-MAR-2001 (TREMELrel. 15, Last sequence update)  
DT 01-MAR-2001 (TREMELrel. 23, Last annotation update)  
DE Genomic DNA, chromosome 5, Pl clone:MRA19.  
OS Arabidopsis thaliana (Mouse-ear cross).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Columbia;  
RX MEDLINE=98403884; PubMed=9734815;  
RA Kotani H., Nakamura Y., Sato S., Asamizu E., Kaneko T., Miyajima N.,  
RA Tabata S.;  
RT "Structural analysis of Arabidopsis thaliana chromosome 5. VI.  
RT Physically assigned Pl and TAC clones."  
RL DNA Res. 5:203-216(1998).  
DR EMBL; AB012245; BAB09212.1; -.  
DR InterPro; IPR004274; NIP.  
DR Pfam; PF03031; NIF; 1.  
DR SMART; SM00577; CPDc; 1.  
SQ SEQUENCE 272 AA; 31510 MW; 912D5789A33BDF29 CRC64;

Query Match 93.8%; Score 30; DB 10; Length 272;  
Best Local Similarity 85.7%; Pred. No. 64;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATINKSL 7  
Db 43 ATINKSL 49

# RESULT 3

Q92NH6 Q92NH6 PRELIMINARY; PRT; 312 AA.  
AC Q92NH6;  
DT 01-DEC-2001 (TREMELrel. 19, Created)  
DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)  
DE Putative D-ERYTHROSE-1-phosphate dehydrogenase protein.  
GN ERYC OR R02228 OR SMC01616.  
OS Rhizobium meliloti (Sinorhizobium meliloti).  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Rhizobiaceae; Sinorhizobium.  
OX NCBI\_TaxID=582;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1021;  
RX MEDLINE=21396507; PubMed=11481430;  
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,  
RA Boissard P., Becker A., Bourry M., Cadieu E., Dreano S., Gloux S.,  
RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,  
RA Pohl T., Portetelle D., Puhler A., Purnelle B., Ramsperger U.,  
RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.;  
RT "Analysis of the chromosome sequence of the legume symbiont  
RT Sinorhizobium meliloti strain 1021."  
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).  
DR EMBL; AL591790; CAC46807.1; -.

KW Complete proteome.  
SQ SEQUENCE 312 AA; 35379 MW; AF09564D6C31601D CRC64;  
Query Match 90.6%; Score 29; DB 16; Length 312;  
Best Local Similarity 85.7%; Pred. No. 1.3e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATINKSL 7  
Db 49 ATINKSL 55

# RESULT 4

Q8R9L4 Q8R9L4 PRELIMINARY; PRT; 433 AA.  
AC Q8R9L4;  
DT 01-JUN-2002 (TREMELrel. 21, Created)  
DT 01-JUN-2002 (TREMELrel. 21, Last sequence update)  
DE Cytosine deaminase and related metal-dependent hydrolases.  
DE SSNA OR TTE1593.  
GN Thermoanaerobacter tengcongensis.  
OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;  
OC Thermoanaerobacteriaceae; Thermoanaerobacter.  
OX NCBI\_TaxID=119072;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MB4 / JCM 11007;  
RX MEDLINE=21992816; PubMed=1197336;  
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,  
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,  
RA Tan H., Chen R., Wang J., Yu J., Yang H.;  
RT "A complete sequence of T. tengcongensis genome."  
RL Genome Res. 12:689-700(2002).  
DR EMBL; AB013114; AAM24797.1; -.  
DR InterPro; IPR006680; Amidohydro\_1.  
DR Pfam; PF01979; Amidohydro\_1; 1.  
KW Hydrolase; Complete proteome.  
SQ SEQUENCE 433 AA; 48349 MW; 53PF2F25463AB8C5 CRC64;

Query Match 90.6%; Score 29; DB 16; Length 433;  
Best Local Similarity 85.7%; Pred. No. 1.8e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATINKSL 7  
Db 316 ATINKSL 322

# RESULT 5

Q94477 Q94477 PRELIMINARY; PRT; 1516 AA.  
AC Q94477;  
DT 01-MAY-1999 (TREMELrel. 10, Created)  
DT 01-MAY-1999 (TREMELrel. 10, Last sequence update)  
DE Putative myosin heavy chain.  
GN SPCC1919.10C.  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972h-;  
RA Lyne M., Harris D.E., Murphey L.D., Rajandream M.A., Barrell B.G.;  
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL035075; CAA22641.1; -.  
DR HSPF; P08799; IWND.  
DR InterPro; IPR002710; DIL.  
DR InterPro; IPR000048; IQ\_region.  
DR InterPro; IPR001609; myosin\_head.

```

InterPro; IPR004009; Myosin_N.
Pfam; PF01843; DIL; 1.
Pfam; PF00612; IQ; 5.
Pfam; PF00063; myosin_head; 1.
Pfam; PF02736; Myosin_N; 1.
PRINTS; PR00193; MYOSINHEAVY
ProDom; PD000355; myosin_head; 1.
SMART; SM00015; IQ; 2.
SMART; SM00242; MYSC; 1.
PROSITE; PS50096; IQ; 2.
SEQUENCE 1516 AA; 175154 MW; 0D25F6FB65B19B39 CRC64;

Query Match 90.6%; Score 29; DB 3; Length 1516;
Best Local Similarity 85.7%; Pred. No. 6.1e-02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OQ 1 ATKNSL 7
Db 432 ATKAL 438
|||||

RESULT 6
Q8KQY8 PRELIMINARY; PRT; 126 AA.
AC Q8KQY8;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein.
DE OS Vibrio cholerae.
DE OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
DE OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=22103115; PubMed=12107144;
RA Beaber J.W., Hochhut B., Waldor M.K.;
RT "Genomic and Functional Analyses of SXT, an Integrating Antibiotic
RT Resistance Gene Transfer Element Derived from Vibrio cholerae.";
RL J. Bacteriol 184:4259-4269(2002).
DR EMBL; AY055428; AAU59726.1; -.
DR Hypothetical protein.
SQ SEQUENCE 126 AA; 14790 MW; AE401A90D4B0E56D CRC64;

Query Match 87.5%; Score 28; DB 2; Length 126;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OQ 2 TINKSL 7
Db 62 TINKSL 67
|||||

RESULT 7
Q8N1L4 PRELIMINARY; PRT; 127 AA.
AC Q8N1L4;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein.
DE SPAB2127.05.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OC NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=372h-;
RA Wood V., Rajadream M.A., Barrell B.G., Harris D.;
RA Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.

```



Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATINKS 6  
 DB 232 ATINKS 237

RESULT 10

Q9N4M9 PRELIMINARY; PRT; 289 AA.

AC Q9N4M9; (TRENBLrel. 15, Created)  
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
 DE Hypothetical 33.4 kDa protein.  
 GN ZK616.7.

OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA None;

RT "Genome sequence of the nematode C. elegans: a platform for  
 RT investigating biology. the C. elegans Sequencing Consortium.";  
 RL Science 282:2012-2018(1998).

RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RX STRAIN=Bristol N2;  
 RA Kellen J., Bradshaw-Cordum H.;  
 RT "The sequence of C. elegans cosmid ZK616.";  
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.

RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Waterston R.;  
 RT "Direct Submission.";  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AC006833; AAF60945.1; -  
 DR HSP; F18052; LYFO.  
 DR WormPep; ZK616.7; CE26345.  
 DR InterPro; IPR000387; TYR\_PHOSPHATASE.  
 DR InterPro; IPR000242; TYR\_PP.  
 DR Pfam; PF00102; Y\_PHOSPHATASE; 1.  
 DR PRINTS; PR00700; PRTYPHPTASE.  
 DR SMART; SM00194; PTTP; 1.  
 DR PROSITE; PS00383; TYR\_PHOSPHATASE\_1; 1.  
 DR PROSITE; PS50056; TYR\_PHOSPHATASE\_2; 1.  
 DR PROSITE; PS50055; TYR\_PHOSPHATASE\_PTP; 1.  
 KW Hypothetical protein; Hydrolase.  
 SQ SEQUENCE 289 AA; 33400 MW; FF956E78BBA47BFA CRC64;

Query Match 87.5%; Score 28; DB 5; Length 289;  
 Best Local Similarity 71.4%; Pred. No. 2e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATINKSL 7  
 DB 139 STINKSL 145

RESULT 11

Q69065 PRELIMINARY; PRT; 368 AA.

AC Q69065;  
 DT 01-NOV-1996 (TRENBLrel. 01, Created)  
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)  
 DE DNA-binding protein.  
 GN P41.

OS Human herpesvirus (type 6 / strain GS) (HHV6).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.

OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Betaherpesvirinae; Roseolovirus.  
 OX NCBI\_TaxID=10369;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GS;  
 RX MEDLINE=93286553; PubMed=8389796;  
 RA Agulnick A.D., Thompson J., Iyengar S., Ablashi D.,  
 RA Ricciardi R.;  
 RT "Identification of a DNA binding protein of human herpesvirus 6, a  
 RT putative DNA polymerase stimulatory factor.";  
 RL J. Gen. Virol. 74:1003-1009(1993).  
 DR EMBL; L12003; AAA43855.1; -  
 DR InterPro; IPR004997; Herpes PAP.  
 DR Pfam; PF03325; Herpes\_PAP; 1.  
 KW DNA-binding  
 SQ SEQUENCE 368 AA; 41912 MW; 4A8693B6C363A3EC CRC64;

Query Match 87.5%; Score 28; DB 12; Length 368;  
 Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TINKSL 7  
 DB 36 TINKSL 41

RESULT 12

Q69463 PRELIMINARY; PRT; 368 AA.

AC Q69463;  
 DT 01-NOV-1996 (TRENBLrel. 01, Created)  
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)  
 DE Protein P41.  
 OS Human herpesvirus (type 6 / strain GS) (HHV6).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Betaherpesvirinae; Roseolovirus.  
 OX NCBI\_TaxID=10369;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GS;  
 RX MEDLINE=94167865; PubMed=8122364;  
 RA Zhou Y., Chang C.K., Qian G., Chandran B., Wood C.;  
 RT "Trans-activation of the HIV promoter by a cDNA and its genomic clones  
 RT of human herpesvirus-6.";  
 RL Virology 199:311-322(1994).  
 DR EMBL; U04437; AAA18340.1; -  
 DR InterPro; IPR004997; Herpes PAP.  
 DR Pfam; PF03325; Herpes\_PAP; 1.  
 SQ SEQUENCE 368 AA; 41841 MW; 1A868914C163A1E3 CRC64;

Query Match 87.5%; Score 28; DB 12; Length 368;  
 Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TINKSL 7  
 DB 36 TINKSL 41

RESULT 13

Q8DW81 PRELIMINARY; PRT; 387 AA.

AC Q8DW81;  
 DT 01-MAR-2003 (TRENBLrel. 23, Created)  
 DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)  
 DE Putative integrase.  
 GN SMU.191C.  
 OS Streptococcus mutans.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.

```

OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UAI159 / ATCC 700610 / Serotype C;
RX MEDLINE=22295063; PubMed=12397186;
RA Ajdic D., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Qian Y.,
RA Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UAI59, a cariogenic dental
RT pathogen."; Acad. Sci. U.S.A. 99:14434-14439(2002).
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
DR EMBL; AEO14870; AAN57965.1; -.
KW Complete proteome.
SQ SEQUENCE 387 AA; 45636 MW; B65557901B2C7E82 CRC64;

Query Match 87.5%; Score 28; DB 16; Length 387;
Best Local Similarity 71.4%; Pred. No. 2.7e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATINKSL 7
Db 307 ATVNKAL 313
||:|||||
||:|||||

RESULT 14
Q9PY12 PRELIMINARY; PRT; 388 AA.
AC Q9PY12;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE ICP36-DNA polymerase accessory protein.
OS Human cytomegalovirus.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxID=10359;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92292234; PubMed=1318399;
RA Ertl P.F., Powell K.L.;
RT "Physical and functional interaction of human cytomegalovirus DNA
RT polymerase and its accessory protein (ICP36) expressed in insect
RT cells.";
RL J. Virol. 66:4126-4133(1992).
DR InterPro; IPR004997; Herpes PAP.
DR Pfam; PF03325; Herpes PAP; I.
SQ SEQUENCE 388 AA; 44174 MW; 76D52759751B365E CRC64;

Query Match 87.5%; Score 28; DB 12; Length 388;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TINKSL 7
Db 56 TINKSL 61
|||||
|||||

RESULT 15
O30859 PRELIMINARY; PRT; 396 AA.
AC O30859;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE Tellurite resistance protein.
GN TEPA.
OS Rhodobacter sphaeroides (Rhodospseudomonas sphaeroides).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
OC Rhodobacteraceae; Rhodobacter.
OX NCBI_TaxID=1063;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2.4.1;
RX MEDLINE=98069464; PubMed=9406390;
RA O'Gara J.P., Gomelsky M., Kaplan S.;
RT "Identification and molecular genetic analysis of multiple loci
RT contributing to high-level tellurite resistance in Rhodobacter
RT sphaeroides 2.4.1.";
RL Appl. Environ. Microbiol. 63:4713-4720(1997).
DR EMBL; AF019377; AAB93480.1; -.
SQ SEQUENCE 396 AA; 43865 MW; 78D31330A15BCB48 CRC64;

Query Match 87.5%; Score 28; DB 2; Length 396;
Best Local Similarity 85.7%; Pred. No. 2.8e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATINKSL 7
Db 342 ATINESL 348
|||||
|||||

Search completed: February 11, 2004, 17:09:21
Job time : 27.5833 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 16:51:43 ; Search time 5.16667 Seconds  
(without alignments)  
63.714 Million cell updates/sec

Title: US-09-901-187C-3

Perfect score: 32

Sequence: 1 ATINKSL 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	29	90.6	557	1 APG6 YEAST	Q02948 saccharomyces
2	29	90.6	1025	1 HTRA_FUGRU	Q02611 fugu rubrip
3	29	90.6	1516	1 MYS4_SCHPO	Q04477 schizosacch
4	28	87.5	190	1 HAS3_CHICK	Q57425 gallus gall
5	28	87.5	388	1 VPAP_HSV6G	P27417 human herpe
6	28	87.5	393	1 VPAP_HSV6G	P52439 human herpe
7	28	87.5	519	1 PDAS_HUMAN	Q14554 homo sapien
8	28	87.5	553	1 HAS3_HUMAN	O00219 homo sapien
9	28	87.5	554	1 HAS3_MOUSE	O08650 mus musculu
10	28	87.5	779	1 IF2_LISMO	Q37756 listeria mo
11	28	87.5	782	1 IF2_LISIN	Q52C23 listeria in
12	28	87.5	947	1 BGLS_RUMAL	P15885 ruminooccu
13	27	84.4	125	1 GLN4_METBA	P54806 methanosarc
14	27	84.4	295	1 THTR_RAT	P24329 rattus norv
15	27	84.4	296	1 THTR_BOVIN	P00586 bos taurus
16	27	84.4	415	1 YV06_CAEEL	P48460 caenorhabdi
17	27	84.4	506	1 DH1L_YEAST	P39517 saccharomyces
18	27	84.4	883	1 KLP5_SCHPO	O43433 schizosacch
19	27	84.4	1097	1 KF1D_RAT	Q35787 rattus norv
20	27	84.4	1103	1 KF1C_HUMAN	O43896 homo sapien
21	27	84.4	1176	1 HMDH_PHYBL	Q12649 phycomyces
22	27	84.4	1584	1 U104_CAEEL	P23678 caenorhabdi
23	27	84.4	1690	1 KF1A_HUMAN	Q12756 homo sapien
24	27	84.4	1695	1 KF1A_MOUSE	P33173 mus musculu
25	27	84.4	1816	1 KF1B_MOUSE	Q60575 mus musculu
26	26	81.2	122	1 CYTA_SARPE	P31727 sarcophaga
27	26	81.2	136	1 VF15_VACCP	P29891 vaccinia vi
28	26	81.2	158	1 VF15_VACCC	P21020 vaccinia vi
29	26	81.2	161	1 VF15_VARV	P33873 variola vir
30	26	81.2	246	1 PCNA_THEAC	Q9bjq0 thymoplasma
31	26	81.2	248	1 UBIG_RHIME	Q8y798 bruceella me
32	26	81.2	248	1 UBIG_RHIME	Q92mk1 rhizobium m
33	26	81.2	249	1 UBIG_RHILO	Q98587 rhizobium 1

34	26	81.2	250	1 UBIG_AGR75	Q8ua66 agrobacteri
35	26	81.2	301	1 YTHH_ECOLI	P39359 escherichia
36	26	81.2	350	1 PANE_SCHPO	Q9hdue schizosacch
37	26	81.2	359	1 LPFD_SALTY	P43663 salmonella
38	26	81.2	390	1 FTSZ_MYCPU	Q50318 mycoplasma
39	26	81.2	479	1 ICES8_HUMAN	Q14790 h caspase-8
40	26	81.2	496	1 AGP2_MOUSE	C35608 mus musculu
41	26	81.2	528	1 ASMA4_YEAST	Q05166 saccharomyc
42	26	81.2	539	1 VGLF_SYVS	P04849 simian viru
43	26	81.2	599	1 SYD_SYNY3	P73851 synectocyst
44	26	81.2	667	1 VTER_HSV6U	P24443 human herpe
45	26	81.2	671	1 KF2C_MACFA	Q951p1 macaca fasc

ALIGNMENTS

RESULT 1

APG6 YEAST

ID APG6 YEAST STANDARD; PRT; 557 AA.

AC Q02948;

DT 15-JUL-1999 (Rel. 38, Created)

DT 15-JUL-1999 (Rel. 38, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Autophagy protein APG6.

GN APG6 OR VPS30 OR VPT30 OR YPL120W OR LPH7.

OS Saccharomyces cerevisiae (Baker's Yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

OX NCBI\_TaxID=4932;

RN [1]

SEQUENCE FROM N.A.

RX MEDLINE=98380442; PubMed=9712845;

RA Kametaka S., Okano T., Ohsumi M., Ohsumi Y.;

RT "Apgl4p and Apge6/Vps30p form a protein complex essential for autophagy in the yeast, Saccharomyces cerevisiae.";

RL J. Biol. Chem. 273:22284-22291(1998).

RN [2]

SEQUENCE FROM N.A.

RC STRAIN=S288c / AB972;

RX MEDLINE=97313271; PubMed=9169875;

RA Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Ansoorge W., Araujo R., Aparicio A., Barrell B.G., Badcock K., Benes V., Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M., Chung E., Churcher C.M., Coster F., Davis K., Davis R.W., Dietrich F.S., Delius H., DiPaolo T., Dubois E., Duesterhoeft A., Duncan M., Floeth M., Fortin N., Friese J.D., Fritz C., Goffeau A., Hall J., Hebling U., Heumann K., Hilbert H., Hillier L., Hunnicke-Smith S., Hyman R., Johnston M., Kalman S., Klein K., Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J., Marathe R., Messenguy F., Mewes H.-W., Mirtipati S., Moestl D., Mueller-Auer S., Namath A., Nentwich U., Oefner P., Pearson D., Petel F.X., Pohl T.M., Purrelle D., Schafer M., Scharfe M., Scherens B., Schramm S., Schroeder M., Seicu A.M., Tettelin H., Urrestazu L.A., Ushinsky S., Vierendeels F., Vissers S., Voss H., Walsh S.V., Wambutt R., Wang Y., Wedler E., Wedler H., Winnett E., Zhong W.W., Zollner A., Vo D.H., Hani J.;

RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI.";

RL Nature 387:103-105(1997).

CC -1- FUNCTION: REQUIRED FOR AUTOGRAPHY.

CC -1- SUBUNIT: FORMS A COMPLEX WITH APG14.

CC -1- SIMILARITY: BELONGS TO THE BECLIN FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; AB011072; BAA32104.1; -

DR EMBL; U43503; AAB68242.1; -

Query Match 90.6%; Score 29; DB 1; Length 1025;  
Best Local Similarity 85.7%; Pred. No. 52;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATINKSL 7  
DB 615 ATINKNL 621

RESULT 3  
MYS4 SCHPO STANDARD; PRT; 1516 AA.  
ID O94477; P78899; Q9US73;  
AC 15-SEP-2003 (Rel. 42, Created)  
DT 15-SEP-2003 (Rel. 42, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Probable myosin heavy chain C1919.10c.  
GN SPCC1919.10c.  
OS Schizosaccharomyces pombe (Fission Yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
ON NCBI\_TaxID=4896;  
RX [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972;  
RA MEDLINE=21848401; PubMed=11859360;  
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
RA Brooks J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,  
RA Gents S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
RA Holtroyd S., Hornby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
RA Rutherford K., Rutter S., Saunders R., Squares S., Stevens K.,  
RA Skellton J., Simmonds M., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
RA Woodward J., Volckaert G., Aert R., Robben J., Grymoprez B.,  
RA Welter J., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,  
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,  
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,  
RA Daga R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
RA Cerrutti L., Lowe T., McCombie M.R., Paulsen I., Potashkin J.,  
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,  
RT "The genome sequence of Schizosaccharomyces pombe."  
RL Nature 415:871-880(2002).  
[2]  
RP SEQUENCE OF 1001-1516 FROM N.A.  
RC STRAIN=PR745;  
RX MEDLINE=98182722; PubMed=9501991;  
RA Yoshida S., Kato K., Nakai K., Okayama H., Nojima H.;  
RT "Identification of open reading frames in Schizosaccharomyces pombe  
CDNAS."  
RL DNA Res. 4:363-369(1997).  
[3]  
RP SEQUENCE OF 1280-1383 FROM N.A., AND SUBCELLULAR LOCATION.  
RC STRAIN=968 t90;  
RX MEDLINE=20232868; PubMed=10759889;  
RA Ding D.-Q., Tomita Y., Yamamoto A., Chikashige Y., Haraguchi T.,  
RA Hiraoka Y.;  
RT "Large-scale screening of intracellular protein localization in living  
fission yeast cells by the use of a GFP-fusion genomic DNA library."  
RL Genes Cells 5:169-190(2000).  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic; localised at the cell poles and  
septum.  
CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.

DR PIR; S62002; S62002.  
DR SGD; S0006041; VPS30.  
DR GO; GO:0005624; C:membrane fraction; IDA.  
DR GO; GO:0006914; P:autophagy; IMP.  
DR GO; GO:0006623; P:protein-vacuolar targeting; IMP.  
DR Pfam; PF04111; APG6; 1.  
DR Autophagy; Coiled coil.  
FT DOMAIN 189 322  
SQ SEQUENCE 557 AA; 63260 MW; 4BA74B614CABE1C0 CRC64;

Query Match 90.6%; Score 29; DB 1; Length 557;  
Best Local Similarity 85.7%; Pred. No. 28;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATINKSL 7  
DB 368 ATINKNL 374

RESULT 2  
HIRA\_FUGRU STANDARD; PRT; 1025 AA.  
ID O42611;  
AC 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE HIRA protein (TUP1 like enhancer of split protein 1).  
GN HIRA OR TUPLE1.  
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Acanthomorphi; Acanthopterygii; Percormorpha; Tetraodontiformes;  
OC Tetraodontidae; Tetraodontidae; Takifugu.  
ON NCBI\_TaxID=31033;  
RX [1]  
RP SEQUENCE FROM N.A.  
RC MEDLINE=98201624; PubMed=95242981;  
RA Llevadot R., Estivill X., Scambler P., Pritchard M.;  
RT "Isolation and genomic characterization of the TUPLE1/HIRA gene of  
the pufferfish Fugu rubripes."  
RL Gene 207:279-283(1998).  
CC -!- FUNCTION: Could have a part in mechanisms of transcriptional  
regulation similar to that played by yeast Hir1 and Hir2 together.  
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).  
CC -!- SIMILARITY: Contains 4 WD repeats.  
CC -!- SIMILARITY: BELONGS TO THE HIR1 FAMILY OF WD-REPEAT PROTEINS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; U94325; AAC60370.1; -  
DR EMBL; U94324; AAC60369.1; -  
DR InterPro; IPR001680; WD40.  
DR Pfam; PF00400; WD40; 7.  
DR ProDom; PD000018; WD40; 2.  
DR SMART; SM00320; WD40; 8.  
DR PROSITE; PS00678; WD REPEATS 1; 1.  
DR PROSITE; PS00682; WD REPEATS 2; 3.  
DR PROSITE; PS00294; WD REPEATS REGION; 1.  
KW Transcription regulation; Repeat; WD repeat; Nuclear protein.  
FT REPEAT 68 98  
FT REPEAT 129 159 WD 2.  
FT REPEAT 172 202 WD 3.  
FT REPEAT 266 313 WD 4.  
FT DOMAIN 673 682 POLY-ALA.  
FT DOMAIN 685 688 POLY-ALA.  
SQ SEQUENCE 1025 AA; 111856 MW; A4212152D75B6A37 CRC64;

CC -!- SIMILARITY: Contains 1 dilute domain.  
 CC -!- SIMILARITY: Contains 5 IQ domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AL035075; CAA22641.1; -  
 CC EMBL; D89250; BA113911.1; ALT\_SEQ.  
 CC EMBL; AB028001; BA87305.1; -  
 CC HSPF; P08799; MNM.  
 CC GenDB; Srombe; SPC1919.10c; -  
 CC InterPro; IPR002710; DIL.  
 CC InterPro; IPR000048; IQ region.  
 CC InterPro; IPR001609; myosin head.  
 CC InterPro; IPR004009; myosin\_N.  
 CC Pfam; PF01843; DIL; 1.  
 CC Pfam; PF00612; IQ; 5.  
 CC Pfam; PF00063; myosin head; 1.  
 CC Pfam; PF02736; Myosin\_N; 1.  
 CC PRINTS; PR00193; MYOSINHEAVY.  
 CC ProDom; PD000355; myosin\_head; 1.  
 CC SMART; SM00015; IQ; 2.  
 CC SMART; SM00242; MYSC; 1.  
 CC PROSITE; PS50096; IQ; 2.  
 CC Myosin; ATP-binding; Actin-binding; Calmodulin-binding; Repeat;  
 KW Coiled coil; Alkylation.  
 FT DOMAIN 75 754 MYOSIN HEAD-LIKE.  
 FT DOMAIN 793 813 IQ 1.  
 FT DOMAIN 818 838 IQ 2.  
 FT DOMAIN 840 865 IQ 3.  
 FT DOMAIN 866 886 IQ 4.  
 FT DOMAIN 888 917 IQ 5.  
 FT DOMAIN 926 1034 COILED COIL (POTENTIAL).  
 FT DOMAIN 1312 1413 DILUTE.  
 FT NP\_BIND 167 174 ATP (POTENTIAL).  
 FT MOD\_RES 688 688 ALKYLATION (BY SIMILARITY).  
 FT MOD\_RES 698 698 ALKYLATION (BY SIMILARITY).  
 FT SEQUENCE 1516 AA; 175154 MW; 0D25P6FB65B19B39 CRC64;  
 QY Query Match 90.6%; Score 29; DB 1; Length 1516;  
 Db Best Local Similarity 85.7%; Pred. No. 79;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ATINSL 7  
 Db 432 ATINSL 438  
 RESULT 4  
 ID HAS3 CHICK STANDARD; PRT; 190 AA.  
 AC OS7425;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Hyaluronan synthase 3 (EC 2.4.1.212) (Hyaluronate synthase 3)  
 DE (Hyaluronic acid synthase 3) (HAS3) (Fragment).  
 GN HAS3.  
 OS Gallus gallus (Chicken).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 CC Gallus.  
 CC NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN-White leghorn;  
 RX MEDLINE=98113148; PubMed=9442026;  
 RA Spicer A.P., McDonald J.A.;

RT "Characterization and molecular evolution of a vertebrate hyaluronan  
 RT synthase gene family";  
 RL J. Biol. Chem. 273:1923-1932(1998).  
 CC -!- FUNCTION: PLAYS A ROLE IN HYALURONAN/HYALURONIC ACID (HA)  
 CC SYNTHESIS (BY SIMILARITY).  
 CC -!- CATALYTIC ACTIVITY: N UDP-N-acetyl-D-glucosamine + n UDP-D-  
 CC glucuronate = [beta-N-acetyl-D-glucosaminyl(1->4)beta-D-  
 CC glucuronosyl(1->3)](n) + 2n UDP.  
 CC -!- COFACTOR: Magnesium.  
 CC -!- PATHWAY: Hyaluronate synthesis.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).  
 CC -!- SIMILARITY: BELONGS TO THE NODC/HAS FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AF015777; AAB94538.1; -  
 CC Transferase; Glycosyltransferase; Transmembrane; Multigene family.  
 KW NON\_TER 1  
 FT DOMAIN <1 103 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 104 124 3 (POTENTIAL).  
 FT DOMAIN 125 134 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 135 155 4 (POTENTIAL).  
 FT DOMAIN 156 161 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 162 182 5 (POTENTIAL).  
 FT DOMAIN 183 >190 EXTRACELLULAR (POTENTIAL).  
 FT NON\_TER 190 190  
 FT SEQUENCE 190 AA; 22533 MW; 2885495BC09D1B3 CRC64;  
 QY Query Match 87.5%; Score 28; DB 1; Length 190;  
 Db Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ATINKS 6  
 Db 185 ATINKS 190  
 RESULT 5  
 ID VPAP HSV6G STANDARD; PRT; 388 AA.  
 AC P27417;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE DNA polymerase processivity factor (Polymerase accessory protein)  
 DE (PAP) (Phosphoprotein P41) (PP41) (Fragment).  
 GN U27.  
 OS Human herpesvirus (type 6 / strain GS) (HHV6).  
 CC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 CC Betaherpesvirinae; Roseolovirus.  
 CC NCBI\_TaxID=10369;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91237802; PubMed=1851860;  
 RX Chang C.K., Balachandran N.;  
 RT Identification, characterization, and sequence analysis of a cDNA  
 RT encoding a phosphoprotein of human herpesvirus 6.;  
 RL J. Virol. 65:2884-2894(1991).  
 RN [2]  
 RP REVISIONS.  
 RA Chang C.K., Balachandran N.;  
 RA Submitted (XX-1994) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: ACCESSORY SUBUNIT OF THE DNA POLYMERASE THAT ACTS TO  
 CC INCREASE THE PROCESSIVITY OF POLYMERIZATION (BY SIMILARITY).  
 CC -!- SIMILARITY: TO OTHER HERPESVIRUSES POLYMERASE ACCESSORY PROTEIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 CC EMBL; M62700; AAA43853.1; ALT INIT.  
 CC InterPro; IPR004997; Herpes\_PAP.  
 CC Pfam; PF03325; Herpes\_PAP; I.  
 CC DNA-binding; DNA replication; Phosphorylation.  
 FT NON\_TER 1  
 SQ SEQUENCE 388 AA; 44184 MW; 2B49595B6414899E5 CRC64;

Query Match 87.5%; Score 28; DB 1; Length 388;  
 Best Local Similarity 100.0%; Pred. No. 33;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TINKSL 7

Db 56 TINKSL 61

# RESULT 6

VPAP\_HSV6U STANDARD; PRT; 393 AA.  
 AC PS2439;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE DNA polymerase processivity factor (Polymerase accessory protein)  
 DE (PAP) (Phosphoprotein P41) (PP41).  
 GN U27 OR EPLF1.  
 OS Human herpesvirus (type 6 / strain Uganda-1102) (HHV6).  
 CC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 CC Betaherpesvirinae; Roseolovirus.  
 CC NCBI\_TaxID=10370;  
 RN [1]  
 RP MEDLINE=95266321; PubMed=7747482;  
 RA Gompels U.A., Nicholas J., Lawrence G., Jones M., Thomson B.J.,  
 RA Martin M.E., Efsthathiou S., Craxton M., Macaulay H.A.;  
 RT "The DNA sequence of human herpesvirus-6: structure, coding content,  
 RT and genome evolution";  
 RL Virology 209:29-51(1995).

CC -1- FUNCTION: ACCESSORY SUBUNIT OF THE DNA POLYMERASE THAT ACTS TO  
 CC INCREASE THE PROCESSIVITY OF POLYMERIZATION (BY SIMILARITY).  
 CC -1- SIMILARITY: TO OTHER HERPESVIRUSES POLYMERASE ACCESSORY PROTEIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 CC EMBL; X83413; CAAS8407.1; -  
 CC InterPro; IPR004997; Herpes\_PAP.  
 CC Pfam; PF03325; Herpes\_PAP; I.  
 CC DNA-binding; DNA replication; Phosphorylation.  
 KW DNA-binding; DNA replication; Phosphorylation.  
 SQ SEQUENCE 393 AA; 44810 MW; 239ADF63F645D90 CRC64;

Query Match 87.5%; Score 28; DB 1; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 34;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TINKSL 7

Db 61 TINKSL 66

# RESULT 7

## PDAS\_HUMAN

ID PDAS\_HUMAN STANDARD; PRT; 519 AA.  
 AC Q14554;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Protein disulfide isomerase A5 precursor (EC 5.3.4.1) (Protein  
 DE disulfide isomerase-related protein).  
 GN PDIA5 OR PDIR.  
 OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 CC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96000209; PubMed=7556671;  
 RA Hayano T., Kikuchi M.;  
 RT "Molecular cloning of the cDNA encoding a novel protein disulfide  
 RT isomerase-related protein (PDIR).";  
 RL FEBS Lett. 372:210-214(1995).  
 CC -1- CATALYTIC ACTIVITY: Rearrangement of both intrachain and  
 CC interchain disulfide bonds in proteins to form the native  
 CC structures.  
 CC -1- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen (By similarity).  
 CC -1- SIMILARITY: Contains 2 thioredoxin domains.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 CC EMBL; D49490; BAA08451.1; -  
 CC PIR; S66673; S66673.  
 CC HSSP; P07237; IMEX.

CC GO; GO:0005783; C:endoplasmic reticulum; TAS.  
 CC GO; GO:0005489; F:electron transporter activity; TAS.  
 CC GO; GO:0003756; P:protein disulfide isomerase activity; TAS.

CC GO; GO:0006457; P:protein folding; TAS.  
 CC GO; GO:0006950; P:response to stress; TAS.  
 CC InterPro; IPR000886; ER target.  
 CC InterPro; IPR008662; Thiored.

CC InterPro; IPR006663; Thioredox\_dom2.

CC Pfam; PF00085; thiored; 3.

CC PRINTS; PR00421; THIOREDOXIN.

CC PROSITE; PS00014; ER TARGET; 1.

CC PROSITE; PS00194; THIOREDOXIN; 2.

CC Redox-active center; Isomerase; Endoplasmic reticulum; Repeat; Signal.

CC SIGNAL 1 21 POTENTIAL

CC CHAIN 22 519 PROTEIN DISULFIDE ISOMERASE A5.

CC DISULFID 182 185 REDOX-ACTIVE (BY SIMILARITY).

CC DISULFID 305 308 REDOX-ACTIVE (BY SIMILARITY).

CC DISULFID 426 429 REDOX-ACTIVE (BY SIMILARITY).

CC SITE 516 519 PREVENT SEQUENCE FROM ER (POTENTIAL).

CC SEQUENCE 519 AA; 59594 MW; 6083FEB8C019658 CRC64;

Query Match 87.5%; Score 28; DB 1; Length 519;  
 Best Local Similarity 71.4%; Pred. No. 45;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATINKSL 7

Db 337 ATINKAL 343

## RESULT 8

### HAS3\_HUMAN

ID HAS3\_HUMAN STANDARD; PRT; 553 AA.  
 AC O00219; Q9NYPO;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)

28-FEB-2003 (Rel. 41, Last annotation update)  
Hyaluronan synthase 3 (EC 2.4.1.212) (Hyaluronate synthase 3)  
(Hyaluronic acid synthase 3) (HA synthase 3)  
HAS3.  
Homo sapiens (Human).  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
NCBI\_TaxID=9606;  
[1]  
SEQUENCE FROM N.A.  
Spicer A.P.;  
"Molecular characterization of hyaluronan synthase 3.";  
Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
[2]  
SEQUENCE OF 275-464 FROM N.A.  
MEDLINE=97238820; PubMed=9083017;  
Spicer A.P., Olson J.S., McDonald J.A.;  
"Molecular cloning and characterization of a cDNA encoding the third  
putative mammalian hyaluronan synthase.";  
J. Biol. Chem. 272:8957-8961(1997).  
-!- FUNCTION: PLAYS A ROLE IN HYALURONAN/HYALURONIC ACID (HA)  
SYNTHESIS.  
-!- CATALYTIC ACTIVITY: N UDP-N-acetyl-D-glucosamine + n UDP-D-  
glucuronate = [beta-N-acetyl-D-glucosaminyl(1->4)beta-D-  
glucuronosyl(1->3)](n) + 2n UDP.  
-!- COFACTOR: Magnesium.  
-!- PATHWAY: Hyaluronate synthesis.  
-!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).  
-!- SIMILARITY: BELONGS TO THE NODC/HAS FAMILY.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; AF232772; AAF36984.1; -  
EMBL; U86409; AAC51209.1; -  
Genew; HGNC:4820; HAS3.  
DR MIM; 602428; -  
DR GO; GO:0005887; C: integral to plasma membrane; TAS.  
DR GO; GO:0005975; P: carbohydrate metabolism; TAS.  
DR InterPro; IPR001173; Glyco trans 2.  
DR Pfam; PF00535; Glycosyltransferase; 1.  
KW Transferase; Glycosyltransferase; Transmembrane; Multigene family.  
FT DOMAIN 1 15 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 16 36 1 (POTENTIAL).  
FT DOMAIN 37 44 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 45 65 2 (POTENTIAL).  
FT DOMAIN 66 377 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 378 398 3 (POTENTIAL).  
FT DOMAIN 399 408 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 409 429 4 (POTENTIAL).  
FT DOMAIN 430 440 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 441 461 5 (POTENTIAL).  
FT DOMAIN 462 473 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 474 494 6 (POTENTIAL).  
FT DOMAIN 495 515 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 516 536 7 (POTENTIAL).  
FT DOMAIN 537 553 EXTRACELLULAR (POTENTIAL).  
FT CARBOHYD 462 462 N-LINKED (GLCNAC...) (POTENTIAL).  
SQ SEQUENCE 553 AA; 63070 MW; 8193B0C930C9E9A1 CRC64;

Query Match 87.5%; Score 28; DB 1; Length 553;  
Best Local Similarity 100.0%; Pred. No. 48;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATINKS 6  
Db 459 ATINKS 464

RESULT 9  
HAS3 MOUSE STANDARD; PRT; 554 AA.  
ID HAS3 MOUSE  
AC O08650;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Hyaluronan synthase 3 (EC 2.4.1.212) (Hyaluronate synthase 3)  
DE (Hyaluronic acid synthase 3) (HA synthase 3).  
GN HAS3.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RX [1]  
SEQUENCE FROM N.A.  
STRAN=CS7BL/6, and 129/SvJ;  
MEDLINE=97238820; PubMed=9083017;  
Spicer A.P., Olson J.S., McDonald J.A.;  
"Molecular cloning and characterization of a cDNA encoding the third  
putative mammalian hyaluronan synthase.";  
J. Biol. Chem. 272:8957-8961(1997).  
-!- FUNCTION: PLAYS A ROLE IN HYALURONAN/HYALURONIC ACID (HA)  
SYNTHESIS.  
-!- CATALYTIC ACTIVITY: N UDP-N-acetyl-D-glucosamine + n UDP-D-  
glucuronate = [beta-N-acetyl-D-glucosaminyl(1->4)beta-D-  
glucuronosyl(1->3)](n) + 2n UDP.  
-!- COFACTOR: Magnesium.  
-!- PATHWAY: Hyaluronate synthesis.  
-!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).  
-!- DEVELOPMENTAL STAGE: EXPRESSED AT E17.5.  
-!- SIMILARITY: BELONGS TO THE NODC/HAS FAMILY.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; U86408; AAC33128.1; -  
MGI; MGI:109599; Has3.  
DR InterPro; IPR001173; Glyco trans 2.  
DR Pfam; PF00535; Glycosyltransferase; 1.  
KW Transferase; Glycosyltransferase; Transmembrane; Multigene family.  
FT DOMAIN 1 15 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 16 36 1 (POTENTIAL).  
FT DOMAIN 37 44 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 45 65 2 (POTENTIAL).  
FT DOMAIN 66 378 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 379 399 3 (POTENTIAL).  
FT DOMAIN 400 409 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 410 430 4 (POTENTIAL).  
FT DOMAIN 431 441 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 442 462 5 (POTENTIAL).  
FT DOMAIN 463 474 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 475 495 6 (POTENTIAL).  
FT DOMAIN 496 516 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 517 537 7 (POTENTIAL).  
FT DOMAIN 538 554 EXTRACELLULAR (POTENTIAL).  
FT CARBOHYD 463 463 N-LINKED (GLCNAC...) (POTENTIAL).  
SQ SEQUENCE 554 AA; 63338 MW; 88296DCB465CE3EC CRC64;

Query Match 87.5%; Score 28; DB 1; Length 554;  
Best Local Similarity 100.0%; Pred. No. 48;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATINKS 6  
Db 460 ATINKS 465

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATKNSL 7  
 DB 228 ATINQSL 234

RESULT 10  
 ID IF2 LISMO STANDARD; PRT; 779 AA.  
 AC 08Y7E5;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Translation initiation factor IF-2.  
 GN INFB OR LMO1325.  
 OS Listeria monocytogenes.  
 OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.  
 OX NCBI\_TaxID=1639;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=EGD-e / Serovar 1/2a;  
 RX MEDLINE=21537279; PubMed=11679669;  
 RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,  
 RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,  
 RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,  
 RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,  
 RA Etian K.-D., Eishi H., Garcia-del Portillo F., Garrido P.,  
 RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,  
 RA Jones L.-M., Kaerst U., Kref J., Kuhn M., Kunst F., Kurapkat G.,  
 RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,  
 RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,  
 RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,  
 RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.,  
 RA "Comparative genomics of Listeria species."  
 RL Science 294:849-853(2001).  
 CC -1- FUNCTION: One of the essential components for the initiation of  
 CC protein synthesis. Protects formylmethionyl-tRNA from spontaneous  
 CC hydrolysis and promotes its binding to the 30S ribosomal subunits.  
 CC Also involved in the hydrolysis of GTP during the formation of the  
 CC 70S ribosomal complex (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: BELONGS TO THE IF-2 FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AL591978; CAC99403.1; -  
 CC PIR; AE1240; AE1240.  
 CC Listlist; LMO01325; -  
 CC HAMAP; MF\_00100; -; 1.  
 CC InterPro; IPR000795; EF\_GTPbind.  
 CC InterPro; IPR004161; EFTU\_D2.  
 CC InterPro; IPR000178; IF2.  
 CC InterPro; IPR006847; IF2\_N.  
 CC InterPro; IPR005225; Small\_GTP.  
 CC Pfam; PF03144; GTP\_EFTU\_D2; 2.  
 CC Pfam; PF04760; IF2\_N; 2.  
 CC PRINTS; PR00315; ELONGATNFCT.  
 CC PRODOM; PD186100; IF2; 1.  
 CC TIGRFAMs; TIGR00487; IF-2; 1.  
 CC TIGRFAMs; TIGR00231; small\_GTP; 1.  
 CC PROSITE; PS01176; IF2; 1.  
 CC Initiation factor; Protein biosynthesis; GTP-binding;  
 CC Complete proteome.  
 CC DOMAIN 283 431 G-DOMAIN  
 CC NP\_BIND 289 296 GTP (BY SIMILARITY).  
 CC FT NP\_BIND 335 339 GTP (BY SIMILARITY).  
 CC FT NP\_BIND 389 392 GTP (BY SIMILARITY).  
 CC SQ SEQUENCE 779 AA; 84576 MW; B377A8DC28294AFF CRC64;  
 Query Match 87.5%; Score 28; DB 1; Length 779;  
 Best Local Similarity 85.7%; Pred. No. 69;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATKNSL 7  
 DB 228 ATINQSL 234

RESULT 11  
 ID IF2 LISIN STANDARD; PRT; 782 AA.  
 AC 092C29;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Translation initiation factor IF-2.  
 GN INFB OR LIN1362.  
 OS Listeria innocua.  
 OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.  
 OX NCBI\_TaxID=1642;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CLIP 11262 / Serovar 6a;  
 RX MEDLINE=21537279; PubMed=11679669;  
 RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,  
 RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,  
 RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,  
 RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,  
 RA Entian K.-D., Eishi H., Garcia-del Portillo F., Garrido P.,  
 RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,  
 RA Jones L.-M., Kaerst U., Kref J., Kuhn M., Kunst F., Kurapkat G.,  
 RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,  
 RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,  
 RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,  
 RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.,  
 RA "Comparative genomics of Listeria species."  
 RL Science 294:849-853(2001).  
 CC -1- FUNCTION: One of the essential components for the initiation of  
 CC protein synthesis. Protects formylmethionyl-tRNA from spontaneous  
 CC hydrolysis and promotes its binding to the 30S ribosomal subunits.  
 CC Also involved in the hydrolysis of GTP during the formation of the  
 CC 70S ribosomal complex (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: BELONGS TO THE IF-2 FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AL596169; CAC96593.1; -  
 CC PIR; A11602; A11602.  
 CC Listlist; LIN01362; -  
 CC HAMAP; MF\_00100; -; 1.  
 CC InterPro; IPR000795; EF\_GTPbind.  
 CC InterPro; IPR004161; EFTU\_D2.  
 CC InterPro; IPR000178; IF2.  
 CC InterPro; IPR006847; IF2\_N.  
 CC InterPro; IPR005225; Small\_GTP.  
 CC Pfam; PF03144; GTP\_EFTU\_D2; 2.  
 CC Pfam; PF04760; IF2\_N; 2.  
 CC PRINTS; PR00315; ELONGATNFCT.  
 CC PRODOM; PD186100; IF2; 1.  
 CC TIGRFAMs; TIGR00487; IF-2; 1.  
 CC TIGRFAMs; TIGR00231; small\_GTP; 1.  
 CC PROSITE; PS01176; IF2; 1.  
 CC Initiation factor; Protein biosynthesis; GTP-binding;  
 CC Complete proteome.  
 CC DOMAIN 286 434 G-DOMAIN  
 CC NP\_BIND 292 299 GTP (BY SIMILARITY).  
 CC FT NP\_BIND 344 351 GTP (BY SIMILARITY).



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 16:52:34 ; Search time 24.5833 Seconds  
(without alignments)  
73.479 Million cell updates/sec

Title: US-09-901-187C-1

Perfect score: 42

Sequence: 1 WQTRKD 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL.23.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_thc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirs.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37	88.1	450	5 P91042	P91042 caenorhabdi
2	37	88.1	555	5 Q95Y99	Q95Y99 caenorhabdi
3	36	85.7	745	16 Q9HU94	Q9HU94 pseudomonas
4	35	83.3	91	10 Q9STRO	Q9STRO arabidopsis
5	35	83.3	104	10 Q9M0J9	Q9M0J9 arabidopsis
6	35	83.3	198	10 Q8XC0	Q8XC0 arabidopsis
7	34	81.0	339	16 Q92E70	Q92E70 rhizobium m
8	34	81.0	747	11 Q8BZA5	Q8BZA5 mus musculu
9	34	81.0	783	4 Q96NR3	Q96NR3 homo sapien
10	33	78.6	147	4 Q8NBE9	Q8NBE9 homo sapien
11	33	78.6	194	11 Q99MS3	Q99MS3 mus musculu
12	33	78.6	205	16 Q8UF59	Q8UF59 agrobacteri
13	33	78.6	225	2 Q8KLR1	Q8KLR1 staphylococ
14	33	78.6	247	2 Q30928	Q30928 streptococ
15	33	78.6	257	5 Q21578	Q21578 caenorhabdi
16	33	78.6	265	16 Q915Y6	Q915Y6 pseudomonas

17	33	78.6	281	2 Q53749	Q53749 staphylococ
18	33	78.6	307	2 Q87370	Q87370 staphylococ
19	33	78.6	307	2 Q8KQF6	Q8KQF6 staphylococ
20	33	78.6	364	4 Q8IUN0	Q8IUN0 homo sapien
21	33	78.6	372	16 Q53860	Q53860 mycobacteri
22	33	78.6	461	2 Q9RH61	Q9RH61 bradyrhizob
23	33	78.6	472	4 Q9Y592	Q9Y592 homo sapien
24	33	78.6	529	10 Q8W0B4	Q8W0B4 oryza sativ
25	33	78.6	637	16 Q9KSG5	Q9KSG5 vibrio chol
26	33	78.6	677	10 Q42426	Q42426 arabidopsis
27	33	78.6	677	10 Q39128	Q39128 arabidopsis
28	33	78.6	765	10 Q942P0	Q942P0 oryza sativ
29	33	78.6	1482	5 Q9V4Y0	Q9V4Y0 drosophila
30	32	76.2	161	13 Q90W12	Q90W12 brachydanio
31	32	76.2	170	3 Q9HDP4	Q9HDP4 zygocacchar
32	32	76.2	221	5 Q19601	Q19601 caenorhabdi
33	32	76.2	250	4 Q8TBL1	Q8TBL1 homo sapien
34	32	76.2	273	11 Q8CAG2	Q8CAG2 mus musculu
35	32	76.2	281	17 Q8ZUQ7	Q8ZUQ7 pyrobaculum
36	32	76.2	287	4 Q96CD8	Q96CD8 homo sapien
37	32	76.2	319	6 Q18787	Q18787 bos taurus
38	32	76.2	319	16 Q8YVC4	Q8YVC4 anabaena sp
39	32	76.2	322	16 Q67737	Q67737 aquifex aeo
40	32	76.2	327	5 Q8T8T4	Q8T8T4 drosophila
41	32	76.2	327	5 Q9VUD8	Q9VUD8 drosophila
42	32	76.2	333	5 Q9BXN9	Q9BXN9 caenorhabdi
43	32	76.2	348	5 Q966U2	Q966U2 moina macro
44	32	76.2	348	5 Q9XTL1	Q9XTL1 daphnia mag
45	32	76.2	348	5 Q95NKK8	Q95NKK8 moina macro

#### ALIGNMENTS

RESULT 1  
P91042 PRELIMINARY; PRT; 450 AA.  
AC P91042;  
DT 01-MAY-1997 (TREMBLrel. 03, Created)  
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)  
DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE Hypothetical 51.3 kDa protein.  
GN C10H11.10.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae; Rhabditidae;  
OC Rhabditidae; Peleoderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
EX MEDLINE=93069613; PubMed=9851916;  
RA None;  
RT "Genome sequence of the nematode C. elegans: a platform for  
investigating biology. The C. elegans Sequencing Consortium.";  
RL Science 282:2012-2018(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Dante M., Wamley P.;  
RT "The sequence of C. elegans cosmid C10H11.";  
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Waterston R.;  
RT "Direct Submission";  
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U88311; AAB42349.1; -;  
DR WormPep; C10H11.10; CE08099.  
KW Hypothetical protein.  
SQ SEQUENCE 450 AA; 51300 MW; 164C00E94F0501A5 CRC64;  
Query Match 88.1%; Score 37; DB 5; Length 450;

Best Local Similarity 85.7%; Pred. No. 36;  
Matches 6; Conservative 0; Mismatches 0; Gaps 0;  
Indels 1; Indels 0; Gaps 0;  
QY 1 WRQTRKD 7  
Db 108 WRQTEKD 114

RESULT 2  
ID Q95Y99 PRELIMINARY; PRT; 555 AA.  
AC Q95Y99;  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE Hypothetical 64.2 kDa protein.  
GN M04F3.3.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1] -  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RX MEDLINE=99069613; PubMed=9851916;  
RA None;  
RT "Genome sequence of the nematode C. elegans: a platform for  
investigating biology. The C. elegans Sequencing Consortium.";  
RL Science 282:2012-2018 (1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Dante M., Kramer J.;  
RT "The sequence of C. elegans cosmid M04F3.3";  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Waterston R.;  
RT "Direct Submission";  
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC006676; AAK71382.1; -  
DR WormPep; M04F3.3; CE12416.  
DR InterPro; IPR000719; Prot kinase.  
DR Pfam; PF00069; Prot kinase; 1.  
DR ProDom; PD000001; Prot kinase; 1.  
DR PROSITE; PS00011; PROTEIN KINASE\_DOM; 1.  
DR Hypothetical protein; ATP-binding; Transferase.  
KW SEQUENCE 555 AA; 64231 MW; 7AAB36EC01A8B6A8 CRC64;  
SQ SEQUENCE 555 AA; 64231 MW; 7AAB36EC01A8B6A8 CRC64;

Query Match 88.1%; Score 37; DB 5; Length 555;  
Best Local Similarity 85.7%; Pred. No. 44;  
Matches 6; Conservative 0; Mismatches 0; Gaps 0;  
Indels 1; Indels 0; Gaps 0;  
QY 1 WRQTRKD 7  
Db 478 WRQTEKD 484

RESULT 3  
ID Q9HU94 PRELIMINARY; PRT; 745 AA.  
AC Q9HU94;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)  
DE Hypothetical protein PA5089.  
GN PA5089.  
OS Pseudomonas aeruginosa.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Pseudomonas.  
OX NCBI\_TaxID=287;  
RN [1]

SEQUENCE FROM N.A.  
RC STRAIN=ATCC 15692 / PA01;  
RX MEDLINE=20437337; PubMed=10984043;  
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,  
Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;  
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
opportunistic pathogen";  
RL Nature 406:959-964 (2000).  
DR EMBL; AC004921; AG08474.1; -  
DR InterPro; IPR001736; PLD.  
DR Pfam; PF00614; PLDC; 2.  
DR SMART; SM00155; PLDC; 2.  
DR PROSITE; PS0035; PLD; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 745 AA; 83418 MW; D2CACF3E04985E25 CRC64;

Query Match 85.7%; Score 36; DB 16; Length 745;  
Best Local Similarity 85.7%; Pred. No. 93;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 WRQTRKD 7  
Db 715 WRQTRKD 721

RESULT 4  
ID Q9STRO PRELIMINARY; PRT; 91 AA.  
AC Q9STRO;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
DE Hypothetical 11.0 kDa protein (Fragment).  
GN T25K17.50.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;  
OC Eucosids II; Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Bevan M., Koetter P., Hempel S., Entian K.-D., Bancroft I.,  
Mewes H.W., Mayer K.F.X., Schueller C.;  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA EU Arabidopsis sequencing project;  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL049171; CAB38953.1; -  
KW Hypothetical protein.  
FT NON\_TER 91  
SQ SEQUENCE 91 AA; 11017 MW; 2165C9F868063634 CRC64;

Query Match 83.3%; Score 35; DB 10; Length 91;  
Best Local Similarity 71.4%; Pred. No. 18;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 1 WRQTRKD 7  
Db 8 WRQTRKD 14

RESULT 5  
ID Q9MOJ9 PRELIMINARY; PRT; 104 AA.  
AC Q9MOJ9;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)

DE Hypothetical 12.1 kDa protein.  
GN AT4G26240.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Rose M., Hempel S., Entian K.-D., Mewes H.W., Lemcke K., Mayer K.F.X.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA EU Arabidopsis sequencing project;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL161564; CAB79479.1; -  
KW Hypothetical protein.  
SQ SEQUENCE 104 AA; 12134 MW; B4DC4A92D37D3F14 CRC64;  
Query Match 83.3%; Score 35; DB 10; Length 104;  
Best Local Similarity 71.4%; Pred. No. 20;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 1 WRQTRKD 7  
DB 8 WRRTRRD 14  
RESULT 6  
Q9RXC0 PRELIMINARY; PRT; 198 AA.  
AC Q9RXC0;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Hypothetical protein.  
GN AT4G26240.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,  
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,  
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,  
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,  
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,  
RA Ecker J., Theologis A., Davis R.W.;  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Haas B.J., Volfovsky N., Town C.D., Troupkhan M., Alexandrov N.,  
RA Feldmann K.A., Flavell R.B., White O., Salzberg S.B.;  
RT "Full-length messenger RNA sequences greatly improve genome  
RT annotation."  
RL Genome Biol. 0:0-0(2002).  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Brover V., Troupkhan M., Alexandrov N., Lu Y.-P., Flavell R.,  
RA Feldmann K.;  
RT "Full-length cDNA from Arabidopsis thaliana."  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Tripp M., Southwick A., Karlin-Neumann G., Nguyen M., Miranda M.,  
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,  
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,  
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,  
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,  
RA Ecker J., Theologis A., Davis R.W.;  
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AY081357; AAL91246.1; -  
DR EMBL; AY085291; AAM62523.1; -  
DR EMBL; AV128848; AAM91248.1; -  
KW Hypothetical protein.  
SQ SEQUENCE 198 AA; 23962 MW; BEC8C8367F65A1AE CRC64;  
Query Match 83.3%; Score 35; DB 10; Length 198;  
Best Local Similarity 71.4%; Pred. No. 38;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 1 WRQTRKD 7  
DB 8 WRRTRRD 14  
RESULT 7  
Q92R70 PRELIMINARY; PRT; 339 AA.  
AC Q92R70;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Putative glycosyltransferase transmembrane protein.  
GN R01040 OR SMC02384.  
OS Rhizobium meliloti (Sinorhizobium meliloti).  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Rhizobiaceae; Sinorhizobium.  
OX NCBI\_TaxID=382;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1021;  
RX MEDLINE=21396507; PubMed=11481430;  
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe P., Batut J.,  
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,  
RA Godrie T., Goffeau A., Kahn D., Kiss E., Leilaure V., Masny D.,  
RA Pohl T., Portetle D., Puehler A., Fumelle B., Rameberger O.,  
RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.;  
RT "Analysis of the chromosome sequence of the legume symbiont  
RT Sinorhizobium meliloti strain 1021."  
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).  
DR EMBL; AL591785; CAC45612.1; -  
DR InterPro; IPR001173; Glyco\_transf\_2.  
DR Pfam; PF00535; Glycos\_transf\_2; 1.  
KW Complete proteome.  
SQ SEQUENCE 339 AA; 37729 MW; 2801CFCAB7AF7C27 CRC64;  
Query Match 81.0%; Score 34; DB 16; Length 339;  
Best Local Similarity 71.4%; Pred. No. 1e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 WRQTRKD 7  
DB 138 WRQNRD 144  
RESULT 8  
Q8BZA5 PRELIMINARY; PRT; 747 AA.  
AC Q8BZA5;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE CDNA FLJ30296 FIS (Fragment).  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;  
RX MEDLINE=22354683; PubMed=12466851;  
RA The FANTOM Consortium,  
RA the RIKEN Genome Exploration Research Group Phase I & II Team;

"Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs."  
RL Nature 420:563-573 (2002).  
DR EMBL; AK036111; BAC29308.1; --  
FT NON TER 747  
SQ SEQUENCE 747 AA; 85307 MW; 5D64523F43270008 CRC64;  
  
Query Match 81.0%; Score 34; DB 11; Length 747;  
Best Local Similarity 71.4%; Pred. No. 2.3e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 WRQTRKD 7  
Db 246 WRKTRED 252  
|||||  
  
RESULT 9  
Q96NR3 PRELIMINARY; PRT; 783 AA.  
ID Q96NR3  
AC Q96NR3  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DE Hypothetical protein FLJ30296.  
DE Hypothetical protein FLJ30296.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Cerebellum;  
RA Nishi T., Nakagawa S., Senoh A., Mizuguchi H., Inagaki H.,  
RA Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S.,  
RA Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,  
RA Kimura K., Yamashita H., Matsuo K., Nakamura Y., Sekine M.,  
RA Kikuchi H., Kanda K., Wagatsuma M., Murakawa K., Kanehori K.,  
RA Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B., Suzuki Y.,  
RA Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.,  
RA "NEO" human cDNA sequencing project."  
RT Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
RL EMBL; AK054858; BAB70816.1; --  
DR EMBL; AK054858; BAB70816.1; --  
DR InterPro; IPR003392; Patched.  
DR InterPro; IPR000731; SSD 5TW.  
DR Pfam; PF02460; Patched; 1.  
DR PROSITE; PS0156; SSD; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 783 AA; 89370 MW; F130E37D07FF84B5 CRC64;  
  
Query Match 81.0%; Score 34; DB 4; Length 783;  
Best Local Similarity 71.4%; Pred. No. 2.4e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 WRQTRKD 7  
Db 246 WRKTRED 252  
|||||  
  
RESULT 10  
Q9N8E9 PRELIMINARY; PRT; 147 AA.  
ID Q9N8E9  
AC Q9N8E9  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DE Hypothetical protein FLJ39599.  
DE Hypothetical protein FLJ39599.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC Kawakami B., Sugiyama A., Takemoto M., Sugiyama T., Irie R.,  
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,

RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,  
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,  
RA Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Suzuki Y.,  
RA Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.,  
RT "NEO" human cDNA sequencing project."  
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AK096918; BAC04895.1; --  
KW Hypothetical protein.  
SQ SEQUENCE 147 AA; 16738 MW; AA2FCAC53DB8B597 CRC64;  
  
Query Match 78.6%; Score 33; DB 4; Length 147;  
Best Local Similarity 83.3%; Pred. No. 70;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 WRQTRK 6  
Db 45 WRQTRR 50  
|||||  
  
RESULT 11  
Q99MS3 PRELIMINARY; PRT; 194 AA.  
ID Q99MS3  
AC Q99MS3  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Mpv17-like protein.  
DE Mpv17-like protein.  
GN MPV17L.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6;  
RC MEDLINE=21226134; PubMed11327696;  
RA Iida R., Yasuda T., Tsubota E., Matsuki T., Kishi K.,  
RA Iida R., Yasuda T., Tsubota E., Matsuki T., Kishi K.,  
RT "Cloning, mapping, genomic organization, and expression of mouse M-LP,  
RT a new member of the peroxisomal membrane protein Mpv17 domain  
RT family.".  
RT Biochem. Biophys. Res. Commun. 283:292-296 (2001).  
RL [2]  
RN SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6;  
RC Iida R., Yasuda T., Tsubota E., Matsuki T., Kishi K.,  
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF305634; AAK32113.2; --  
DR MGD; MGI:2135951; Mpv17L.  
SQ SEQUENCE 194 AA; 22180 MW; A0F43AC6F80F6B86 CRC64;  
  
Query Match 78.6%; Score 33; DB 11; Length 194;  
Best Local Similarity 83.3%; Pred. No. 92;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 WRQTRK 6  
Db 45 WRQTRR 50  
|||||  
  
RESULT 12  
Q8UF59 PRELIMINARY; PRT; 205 AA.  
ID Q8UF59  
AC Q8UF59  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Hypothetical protein Atul540.  
DE ATU1540 OR AGR C 2837.  
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Rhizobiaceae; Rhizobium.  
OX NCBI\_TaxID=176299;  
RN [1]

SEQUENCE FROM N.A.  
RX MEDLINE=21608550; PubMed=11743193;  
RA Wood D.W., Satubal J.C., Kaul R., Monks D.E., Kitajima J.P.,  
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,  
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,  
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,  
RA Kutyavina T., Levy R., Li M.-J., McClelland E., Palmieri A.,  
RA Raymond C., Rouse G., Saepthimachak C., Wu Z., Romero P., Gordon D.,  
RA Zhang S., Ioo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,  
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,  
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,  
RA Nester E.W.;  
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens  
C58";  
RL Science 294:2317-2323 (2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21608551; PubMed=11743194;  
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,  
RA Quirillo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,  
RA Houmlel K., Gordon J., Vaudin M., Iatchouk O., Epp A., Liu F.,  
RA Wollam C., Allinger M., Dougherty D., Scott C., Lapras C., Markelz B.,  
RA Flanagan C., Crowell C., Gursion J., Lomo C., Sear C., Strub G.,  
RA Cielo C., Slater S.;  
RT "Genome sequence of the plant pathogen and biotechnology agent  
RT Agrobacterium tumefaciens C58";  
RL Science 294:2323-2328 (2001).  
DR EMBL; AE009112; AAL42545.1; --  
DR EMBL; AE008077; AAK87328.1; --  
KW Hypothetical protein, Complete proteome.  
SQ SEQUENCE 205 AA; 23538 MW; 2362AF9C4D0EC702 CRC64;

Query Match 78.6%; Score 33; DB 16; Length 205;  
Best Local Similarity 83.3%; Pred. No. 98;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WRQTRK 6  
Db 36 WRETRK 41  
||: |||

RESULT 13  
Q8KLR1 PRELIMINARY; PRT; 225 AA.  
ID Q8KLR1  
AC Q8KLR1  
DT 01-OCT-2002 (T-EMBLrel. 22, Created)  
DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)  
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)  
DE Replication protein.  
GN REP.  
OS Staphylococcus aureus.  
CG Plasmid pMC524 /MEM.  
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
OX NCBI\_TaxID=1280;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MC524;  
RA Bal M.;  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ312056; CAD24833.1; --  
DR InterPro; IPR000989; Rep.  
DR Pfam; PF01446; Rep; 1.  
KW Plasmid.  
SQ SEQUENCE 225 AA; 26939 MW; CDA38060A70C5371 CRC64;

Query Match 78.6%; Score 33; DB 2; Length 225;  
Best Local Similarity 71.4%; Pred. NO. 1.1e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 WRQTRK 7  
Db 70 WRKARKD 76  
||: |||

RESULT 14  
O30928 PRELIMINARY; PRT; 247 AA.  
ID O30928  
AC O30928  
DT 01-JAN-1998 (T-EMBLrel. 05, Created)  
DT 01-JAN-1998 (T-EMBLrel. 05, Last sequence update)  
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)  
DE Replication protein.  
GN REP371.  
OS Streptococcus thermophilus.  
CG Plasmid pER371.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1308;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ST371;  
RA Solaiman D.K.Y., Somkuti G.A.;  
RT "Characterization of pER371-based Streptococcus thermophilus-  
RT Escherichia coli shuttle vectors.";  
RL Biotechnol. Lett. 19:595-598 (1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ST371;  
RX MEDLINE=98436178; PubMed=9763687;  
RA Solaiman D.K.Y., Somkuti G.A.;  
RT "Characterization of a novel Streptococcus thermophilus rolling-circle  
RT plasmid used for vector construction.";  
RL Appl. Microbiol. Biotechnol. 50:174-180 (1998).  
DR EMBL; AF022180; AAB80748.1; --  
DR InterPro; IPR000989; Rep.  
DR Pfam; PF01446; Rep; 1.  
KW Plasmid.  
SQ SEQUENCE 247 AA; 29020 MW; 1D940C7A14FCEEE7 CRC64;

Query Match 78.6%; Score 33; DB 2; Length 247;  
Best Local Similarity 71.4%; Pred. No. 1.2e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 WRQTRK 7  
Db 50 WRKARKD 56  
||: |||

RESULT 15  
Q21578 PRELIMINARY; PRT; 257 AA.  
ID Q21578  
AC Q21578  
DT 01-NOV-1996 (T-EMBLrel. 01, Created)  
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)  
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)  
DE Hypothetical 29.7 kDa protein.  
GN M6.3.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RX MEDLINE=99069613; PubMed=9851916;  
RA None;  
RT "Genome sequence of the nematode C. elegans: a platform for  
RT investigating biology. The C. elegans Sequencing Consortium.";  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Minx P., Le T.T.;  
RT "The sequence of C. elegans cosmid M6.";  
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
RN [3]

RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Waterston R.;  
RT "Direct Submission."  
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
CC -|- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
DR EMBL; U56860; AAA98706.1; -;  
DR HSP; P14553; I872.  
DR WormPep; M6.3; CE07384.  
DR InterPro; IPR001356; Homeobox.  
DR InterPro; IPR000047; HTH\_lambrepresr.  
DR Pfam; PF00046; homeobox; 1.  
DR PRINTS; PR00024; HOMEBOX.  
DR PRINTS; PR00031; HTHREPRESSR.  
DR ProDom; PD000010; Homeobox; 1.  
DR SMART; SM00389; HOX; 1.  
DR PROSITE; PS00027; HOMEBOX 1; 1.  
DR PROSITE; PS00071; HOMEBOX 2; 1.  
KW Hypothetical protein; DNA-Binding; Homeobox; Nuclear protein.  
SQ SEQUENCE 257 AA; 29693 MW; D47E959D357392CF CRC64;

Query Match 78.6%; Score 33; DB 5; Length 257;  
Best Local Similarity 71.4%; Pred. No. 1.2e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 WRQTRKD 7  
||: |||  
Db 222 WRRVRKD 228

Search completed: February 11, 2004, 17:09:15  
Job time : 33.5833 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 16:51:43 ; Search time 5.16667 Seconds  
(without alignments)  
63.714 Million cell updates/sec

Title: US-09-901-187C-1

Perfect score: 42

Sequence: 1 WRQTRKD 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36	85.7	332	1	OPPF_HAEIN
2	33	78.6	229	1	REPX_STAAN
3	33	78.6	629	1	ASM_HUMAN
4	33	78.6	1835	1	CCAT_RAT
5	32	76.2	331	1	UL38_HCMVA
6	32	76.2	433	1	CBPA_ANOGA
7	32	76.2	462	1	EF11_CRIGR
8	32	76.2	462	1	EF11_MOUSE
9	32	76.2	462	1	EF1A_CHICK
10	32	76.2	535	1	C7C3_MAIZE
11	32	76.2	614	1	AL11_ARAHY
12	32	76.2	626	1	AL12_ARAHY
13	32	76.2	938	1	TRG1_ECOLI
14	32	76.2	1337	1	PTPJ_HUMAN
15	32	76.2	101	1	PCHB_PSEAE
16	31	73.8	223	1	MOXX_PARDE
17	31	73.8	299	1	NPM_XENLA
18	31	73.8	408	1	YFIN_ECOLI
19	31	73.8	434	1	PXR_HUMAN
20	31	73.8	471	1	CP11_PSEAE
21	31	73.8	474	1	GDB_CAUCR
22	31	73.8	490	1	C7DB_LOTJA
23	31	73.8	496	1	C7B1_THLAR
24	31	73.8	894	1	WTP_HUMAN
25	31	73.8	894	1	CID2_SCHPO
26	30	71.4	167	1	P821_ARATH
27	30	71.4	204	1	CID2_SCHPO
28	30	71.4	226	1	EXOY_RHIME
29	30	71.4	226	1	EXOY_RHISN
30	30	71.4	249	1	REP2_ZYGBA
31	30	71.4	314	1	YK67_YEAST
32	30	71.4	346	1	Y343_MYCGE
33	30	71.4	361	1	YIDS_ECOLI

34	30	71.4	395	1	PVRD_MOUSE
35	30	71.4	395	1	PVRD_RAT
36	30	71.4	487	1	CPCU_CANFA
37	30	71.4	489	1	CPCM_RAT
38	30	71.4	496	1	C7D9_SOYBN
39	30	71.4	500	1	C7D7_SOLCH
40	30	71.4	504	1	C7D8_SOYBN
41	30	71.4	510	1	C7DA_SOYBN
42	30	71.4	520	1	C84A_ARATH
43	30	71.4	522	1	C822_SOYBN
44	30	71.4	544	1	C821_PEA
45	30	71.4	548	1	ACEA_NEUCR

## ALIGNMENTS

### RESULT 1

ID	OPPF_HAEIN	STANDARD;	PRT;	332 AA.
AC	P45051;			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Oligopeptide transport ATP-binding protein oppf.			
GN	OPPF OR H1120.			
OS	Haemophilus influenzae.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;			
OC	Pasteurellaceae; Haemophilus.			
OX	NCBI_TaxID=727;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=rd / KW20 / ATCC 51907;			
RX	MEDLINE=95350630; PubMed=7542800;			
RA	Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,			
RA	Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,			
RA	McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,			
RA	Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,			
RA	Weidman J.F., Phillips C.A., Spriggs T., Hedblom B., Cotton M.D.,			
RA	Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,			
RA	Fine L.D., Fritchman J.D., Fuhrmann J.L., Geoghagan N.S.M.,			
RA	Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,			
RT	Venter J.C.;			
RL	"Whole-genome random sequencing and assembly of Haemophilus influenzae			
RD	Science 269:496-512(1995).			
CC	-!- FUNCTION: PART OF THE BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM			
CC	FOR OLIGOPEPTIDES. PROBABLY RESPONSIBLE FOR ENERGY COUPLING TO THE			
CC	TRANSPORT SYSTEM (BY SIMILARITY).			
CC	-!- SUBCELLULAR LOCATION: Inner membrane-associated (Potential).			
CC	-!- SIMILARITY: Belongs to the ABC transporter family.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
CC	EMBL: U32792; AAC22774.1; -			
DR	FIR; B64184; B64184.			
DR	TIGR; H11120; -			
DR	InterPro: IPR003593; AAA ATPase.			
DR	InterPro: IPR003439; ABC transporter.			
DR	Pfam: PF00005; ABC tran; 1.			
DR	ProDom: PD000006; ABC transporter; 1.			
DR	SMART: SM00382; AAA; 1.			
DR	PROSITE: PS00311; ABC TRANSPORTER 1; 1.			
DR	PROSITE: PS00893; ABC_TRANSPORTER_2; 1.			
KW	Peptide transport; Transport; Inner membrane; ATP-binding;			
KW	Complete proteome.			
FT	NP_BIND 56 63 ATP (POTENTIAL).			

O35435	mus musculus
O63707	rattus norv
P56594	canis fami
P19225	rattus norv
O81971	glycine max
P93531	solanum cha
O81974	glycine max
O48923	glycine max
O42600	arabidopsis
O81972	glycine max
O43068	psium sativ
P28299	neurospora

```

SQ SEQUENCE 332 AA; 36760 MW; 45BGA3085431E3D9 CRC64;
Query Match 85.7%; Score 36; DB 1; Length 332;
Best Local Similarity 71.4%; Pred. No. 4.8;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WRQTRKD 7
Db 95 WKETRKD 101

RESULT 2
REPX STAAU STANDARD; PRT; 229 AA.
AC P03862;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Rep protein (Replication protein) (Reading frame A).
OS Staphylococcus aureus.
OG Plasmid pC194.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1280;
[1]
RN SEQUENCE FROM N.A.; PubMed=6950931;
RX MEDLINE=82167188; PubMed=6950931;
RA Horinouchi S., Weisblum B.;
RT "Nucleotide sequence and functional map of pC194, a plasmid that
RT specifies inducible chloramphenicol resistance.";
RL J. Bacteriol. 150:815-825(1982).
CC -!- FUNCTION: PRODUCES A SINGLE-STRAND NICK IN A SPECIFIC SITE OF THE
CC PLASMID, AND THIS NICK RESULTS IN SINGLE-STRAND REPLICATION BY
CC ROLLING CIRCLE MECHANISM.
CC -!- MISCELLANEOUS: PC194 IS A PLASMID THAT SPECIFIES INDUCIBLE
CC CHLORAMPHENICOL RESISTANCE.
CC -!- SIMILARITY: TO SEVERAL PROTEINS INVOLVED IN GRAM-POSITIVE BACTERIA
CC PLASMIDS REPLICATION.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
-----
DR EMBL; V01277; CAA24595.1; ALT_INIT.
DR PIR; A04491; QQSAEC.
DR InterPro; IPR000989; Rep.
DR Pfam; PF01446; Rep; 1.
KW DNA replication; Plasmid.
FT BINDING 214 DNA (BY SIMILARITY).
SQ SEQUENCE 229 AA; 27546 MW; D0FF587D2BEA18C4 CRC64;
Query Match 78.6%; Score 33; DB 1; Length 229;
Best Local Similarity 71.4%; Pred. No. 13;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 WRQTRKD 7
Db 70 WKARKD 76

RESULT 3
ASM_HUMAN STANDARD; PRT; 629 AA.
ID ASM_HUMAN
AC P17405; P17406; Q13811; Q16837; Q16841;
DT 01-AUG-1990 (Rel. 15, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Sphingomyelin phosphodiesterase precursor (EC 3.1.4.12) (Acid
DE sphingomyelinase) (aSMase).
OX SMPD1 OR ASM.

```

```

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.; AND ALTERNATIVE SPLICING.
RX MEDLINE=91217097; PubMed=1840600;
RA Schuchman E.H., Suchi M., Takahashi T., Sandhoff K., Desnick R.J.;
RT "Human acid sphingomyelinase. Isolation, nucleotide sequence and
RT expression of the full-length and alternatively spliced cDNAs.";
RL J. Biol. Chem. 266:8531-8539(1991).
[2]
RN SEQUENCE FROM N.A.
RX MEDLINE=93183402; PubMed=1292508;
RA Newzella D., Stoffel W.;
RT "Molecular cloning of the acid sphingomyelinase of the mouse and the
RT organization and complete nucleotide sequence of the gene.";
RL Biol. Chem. Hoppe-Seyler 373:1233-1238(1992).
[3]
RN SEQUENCE FROM N.A.
RX MEDLINE=92155708; PubMed=1740330;
RA Schuchman E.H., Levran O., Pereira L.V., Desnick R.J.;
RT "Structural organization and complete nucleotide sequence of the gene
RT encoding human acid sphingomyelinase (SMPD1).";
RL Genomics 12:197-205(1992).
[4]
RN SEQUENCE FROM N.A.; AND VARIANT ARG-157.
RX MEDLINE=94012573; PubMed=8407868;
RA Ida H., Rennett O.M., Eto Y., Chan W.Y.;
RT "Cloning of a human acid sphingomyelinase cDNA with a new mutation
RT that renders the enzyme inactive.";
RL J. Biochem. 114:15-20(1993).
[5]
RN SEQUENCE OF 128-629 FROM N.A., PARTIAL SEQUENCE, AND
RN ALTERNATIVE SPLICING.
RX TISSUE=Fibroblast;
RX MEDLINE=90060003; PubMed=2555181;
RA Quintern L.E., Schuchman E.H., Levran O., Suchi M., Ferlinz K.,
RA Reiske H., Sandhoff K., Desnick R.J.;
RT "Isolation of cDNA clones encoding human acid sphingomyelinase:
RT occurrence of alternatively processed transcripts.";
RL EMBO J. 8:2469-2473(1989).
[6]
RN CARBOHYDRATE-LINKAGE SITES.
RX MEDLINE=97182640; PubMed=9030779;
RA Ferlinz K., Hurwitz R., Moczall H., Lansmann S., Schuchman E.H.,
RA Sandhoff K.;
RT "Functional characterization of the N-glycosylation sites of human
RT acid sphingomyelinase by site-directed mutagenesis.";
RL Eur. J. Biochem. 243:511-517(1997).
[7]
RN DISULFIDE BONDS.
RX MEDLINE=22518502; PubMed=12631268;
RA Lansmann S., Schuette C.G., Bartelsen O., Hoernschemeyer J., Linke T.,
RA Weisgerber J., Sandhoff K.;
RT "Human acid sphingomyelinase.";
RL Eur. J. Biochem. 270:1076-1088(2003).
[8]
RN VARIANT NPA SER-577.
RX MEDLINE=92028849; PubMed=1718266;
RA Ferlinz K., Hurwitz R., Sandhoff K.;
RT "Molecular basis of acid sphingomyelinase deficiency in a patient
RT with Niemann-Pick disease type A.";
RL Biochem. Biophys. Res. Commun. 179:1187-1191(1991).
[9]
RN VARIANT NPA LEU-496.
RX MEDLINE=91219449; PubMed=2023926;
RA Levran O., Desnick R.J., Schuchman E.H.;
RT "Niemann-Pick disease: a frequent missense mutation in the acid
RT sphingomyelinase gene of Ashkenazi Jewish type A and B patients.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:3748-3752(1991).
[10]
RN VARIANT NPB ARG-608 DEL.

```



RX MEDLINE=91358737; PubMed=1885770;  
RX Levran O., Desnick R.J., Schuchman E.H.;  
RA "Niemann-Pick type B disease. Identification of a single codon  
RT deletion in the acid sphingomyelinase gene and genotype/phenotype  
RT correlations in type A and B patients.";  
RL J. Clin. Invest. 88:806-810(1991).  
RN [11]  
RN VARIANT NPA PRO-302.  
RX MEDLINE=93004773; PubMed=1391960;  
RX Levran O., Desnick R.J., Schuchman E.H.;  
RA "Identification and expression of a common missense mutation (L302P)  
RT in the acid sphingomyelinase gene of Ashkenazi Jewish type A  
RT Niemann-Pick disease patients.";  
RL Blood 80:2081-2087(1992).  
RN [12]  
RN VARIANT NPB ARG-436.  
RX MEDLINE=93244834; PubMed=1301192;  
RX Takahashi T., Desnick R.J., Takada G., Schuchman E.H.;  
RA "Identification of a missense mutation (S436R) in the acid  
RT sphingomyelinase gene from a Japanese patient with type B  
RT Niemann-Pick disease.";  
RL Hum. Mutat. 1:70-71(1992).  
RN [13]  
RN VARIANT NPA ILE-382, AND VARIANTS NPB ARG-242 AND SER-383.  
RX MEDLINE=92316934; PubMed=1618760;  
RX Takahashi T., Suchi M., Desnick R.J., Takada G., Schuchman E.H.;  
RA "Identification and expression of five mutations in the human acid  
RT sphingomyelinase gene causing types A and B Niemann-Pick disease.  
RT Molecular evidence for genetic heterogeneity in the neuropathic and  
RT non-neuropathic forms.";  
RL J. Biol. Chem. 267:12552-12558(1992).  
RN [14]  
RN VARIANT NPB GLY-391.  
RX MEDLINE=94328611; PubMed=8051942;  
RX Sperl W., Bart G., Vanier M.T., Christomanou H., Baldissera I.,  
RA Steichenodorf E., Paschke E.;  
RT "A family with visceral course of Niemann-Pick disease, macular halo  
RT syndrome and low sphingomyelin degradation rate.";  
RL J. Inherit. Metab. Dis. 17:93-103(1994).  
RN [15]  
RN VARIANT NPA THR-389.  
RX MEDLINE=96287387; PubMed=8680412;  
RX Schuchman E.H.;  
RA "Two new mutations in the acid sphingomyelinase gene causing type A  
RT Niemann-Pick disease: N389T and R441X.";  
RL Hum. Mutat. 6:352-354(1995).  
RN [16]  
RN VARIANT NPA CYS-446.  
RX MEDLINE=96274768; PubMed=8693491;  
RX Takahashi T., Suchi M., Sato W., Ten S.B., Sakuragawa N.,  
RA Desnick R.J., Schuchman E.H., Takada G.;  
RT "Identification and expression of a missense mutation (Y446C) in the  
RT acid sphingomyelinase gene from a Japanese patient with type A  
RT Niemann-Pick disease.";  
RL Tohoku J. Exp. Med. 177:117-123(1995).  
RN [17]  
RN VARIANT NPA GLN-246.  
RX MEDLINE=96263741; PubMed=8664904;  
RX Ida H., Rennert O.M., Maekawa K., Eto Y.;  
RA "Identification of three novel mutations in the acid  
RT sphingomyelinase gene of Japanese patients with Niemann-Pick disease  
RT type A and B.";  
RL Hum. Mutat. 7:65-67(1996).  
RN [18]  
RN VARIANT NPB TYR-421.  
RX MEDLINE=22340429; PubMed=12369017;  
RX Sinonaro C.M., Desnick R.J., McGovern M.M., Wasserstein M.P.,  
RA Schuchman E.H.;  
RT "The demographics and distribution of type B Niemann-Pick disease:  
RT novel mutations lead to new genotype/phenotype correlations.";  
RL Am. J. Hum. Genet. 71:1413-1419(2002).  
RN [19]  
RN VARIANTS NPA ARG-248; TYR-319; SER-463; LEU-475 AND HIS-537, AND

RP VARIANTS NPB SER-371 AND ARG-608 DEL.  
RX MEDLINE=22444008; PubMed=12556236;  
RX Sikora J., Pavlu-Pereira H., Ellender M., Roelofs H., Wevers R.A.;  
RA "Seven novel Acid sphingomyelinase gene mutations in Niemann-Pick type  
RT A and B patients.";  
RL Ann. Hum. Genet. 67:63-70(2003).  
RN [20]  
RN FUNCTION: CONVERTS SPHINGOMYELIN TO CERAMIDE. ASM ALSO HAS  
CC PHOSPHOLIPASE C ACTIVITIES TOWARD 1,2-DIACYLGLYCEROLPHOSPHOCHOLINE  
CC AND 1,2-DIACYLGLYCEROLPHOSPHOGLYCEROL.  
CC -!- CATALYTIC ACTIVITY: Sphingomyelin + H(2)O = N-acylsphingosine +  
CC choline phosphate.  
CC -!- SUBUNIT: Monomer.  
CC -!- SUBCELLULAR LOCATION: Lysosomal.  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=3;  
CC Name=1; Synonyms=ASM-1;  
CC IsoId=P17405-1; Sequence=Displayed;  
CC Name=2; Synonyms=ASM-2; Sequence=VSP\_000331, VSP\_000332;  
CC IsoId=P17405-2; Sequence=VSP\_000331, VSP\_000332;  
CC Name=3; Synonyms=ASM-3;  
CC IsoId=P17405-3; Sequence=VSP\_000333;  
CC -!- DISEASE: Defects in SMPD1 are the cause of Niemann-Pick disease  
CC type A (NPA) [MIM:257200]; also referred to as the classical  
CC infantile form. Niemann-Pick disease is a clinically and  
CC genetically heterogeneous recessive disorder. It is caused by the  
CC accumulation of sphingomyelin and other metabolically related  
CC lipids in the lysosomes, resulting in neurodegeneration starting  
CC from early life. Patients may show xanthomas, pigmentation,  
CC hepatosplenomegaly, lymphadenopathy and mental retardation.  
CC Niemann-Pick disease occurs more frequently among individuals of  
CC Ashkenazi Jewish ancestry than in the general population. NPA is  
CC characterized by very early onset in infancy and a rapidly  
CC progressive course leading to death by three years.  
CC -!- DISEASE: Defects in SMPD1 are the cause of Niemann-Pick disease  
CC type B (NPB) [MIM:607616]; also referred to as the visceral form.  
CC NPB has little if any neurologic involvement and patients may  
CC survive into adulthood.  
CC -!- MISCELLANEOUS: THERE ARE TWO TYPES OF SPHINGOMYELINASES: ASM  
CC (ACID), AND NSM (NEUTRAL).  
CC -!- MISCELLANEOUS: Isoform 1 is the most abundant (90%), isoforms 2  
CC (10%) and 3 (<1%) are only found at lower levels. Only isoform 1  
CC is a catalytic active enzyme.  
CC -!- SIMILARITY: BELONGS TO THE ACID SPHINGOMYELINASE FAMILY.  
CC -!- SIMILARITY: Contains 1 saposin B-type domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC  
Query Match 78.6%; Score 33; DB 1; Length 629;  
Best Local Similarity 71.4%; Pred. No. 38;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 WQTRKQD 7  
DB 285 WHQTRQD 291  
RESULT 4  
CCAI RAT STANDARD; PRT; 1835 AA.  
ID CCAL RAT  
AC Q9Z0V8; Q9EQ59;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Voltage-dependent T-type calcium channel alpha-1I subunit (CaVt.3).  
GN CACNA1I.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RP STRAIN=Sprague-Dawley; TISSUE=Brain;  
RX MEDLINE=99165796; PubMed=10066244;

Lee J.-H., Daud A.N., Cribbs L.L., Lacerda A.E., Pereverzev A., Klockner U., Schneider T., Perez-Reyes E.; "Cloning and expression of a novel member of the low voltage-activated T-type calcium channel family."; J. Neurosci. 19:1912-1921(1999).

SEQUENCE FROM N.A.

TISSUE=Brain;

MEDLINE=21264893; PubMed=11073957;

McRory J.E., Santi C.M., Hamming K.S.C., Mezeyova J., Sutton K.G., Baillie D.L., Stea A., Snutch T.P.;

"Molecular and functional characterization of a family of rat brain T-type calcium channels.";

J. Biol. Chem. 276:13999-4011(2001).

CC -!- FUNCTION: Voltage-sensitive calcium channels (VSCC) mediate the entry of calcium ions into excitable cells and are also involved in a variety of calcium-dependent processes, including muscle contraction, hormone or neurotransmitter release, gene expression, cell motility, cell division and cell death. Isoform alpha-11 gives rise to T-type calcium currents. T-type calcium channels belong to the "low-voltage activated (LVA)" group and are strongly blocked by nickel and mibefradil. A particularity of this type of channels is an opening at quite negative potentials, and a voltage-dependent inactivation. T-type channels serve pacemaking functions in both central neurons and cardiac nodal cells and support calcium signaling in secretory cells and vascular smooth muscle. They may also be involved in the modulation of firing patterns of neurons which is important for information processing as well as in cell growth processes. Gates in voltage ranges similar to, but higher than alpha 1G or alpha 1H.

CC -!- SUBCELLULAR LOCATION: Integral membrane protein.

CC -!- TISSUE SPECIFICITY: Brain.

CC -!- DOMAIN: Each of the four internal repeats contains five hydrophobic transmembrane segments (S1, S2, S3, S5, S6) and one positively charged transmembrane segment (S4). S4 segments probably represent the voltage-sensor and are characterized by a series of positively charged amino acids at every third position.

CC -!- PTM: In response to raising of intracellular calcium, the T-type channels are activated by Cam-kinase II (By similarity).

CC -!- SIMILARITY: BELONGS TO THE CALCIUM CHANNEL ALPHA-1 SUBUNITS FAMILY.

-----

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

DR EMBL; AF086827; AADI7796.1; -.

DR EMBL; AF290214; AAG35188.1; -.

DR InterPro; IPR001682; Ca/Na\_pore.

DR InterPro; IPR002077; Ca\_channel.

DR InterPro; IPR002111; Cat\_channel\_TrpL.

DR InterPro; IPR005821; Ion\_trans\_nlg.

DR InterPro; IPR005820; M-channel\_nlg.

DR InterPro; IPR005445; TVDCCALPHAL.

DR Pfam; PF00520; ion\_trans; 4.

DR PRINTS; PR01629; CACHANNEL.

DR PRINTS; PR01629; TVDCCALPHAL.

KW Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;

KW Calcium channel; Glycoprotein; Repeat; Multigene family;

KW Calcium-binding; Phosphorylation.

FT REPEAT 64 399 I.

FT REPEAT 584 823 II.

FT REPEAT 1116 1393 III.

FT REPEAT 1431 1692 IV.

FT DOMAIN 1 76 CYTOPLASMIC (POTENTIAL).

FT DOMAIN 77 97 S1 OF REPEAT I (POTENTIAL).

FT DOMAIN 98 115 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 116 137 S2 OF REPEAT I (POTENTIAL).

FT DOMAIN 138 146 CYTOPLASMIC (POTENTIAL).

TRANSMEM	147	166	S3 OF REPEAT I (POTENTIAL).
DOMAIN	167	171	EXTRACELLULAR (POTENTIAL).
TRANSMEM	172	189	S4 OF REPEAT I (POTENTIAL).
DOMAIN	190	209	CYTOPLASMIC (POTENTIAL).
TRANSMEM	210	230	S5 OF REPEAT I (POTENTIAL).
DOMAIN	231	371	EXTRACELLULAR (POTENTIAL).
TRANSMEM	372	396	S6 OF REPEAT I (POTENTIAL).
DOMAIN	397	598	CYTOPLASMIC (POTENTIAL).
TRANSMEM	599	619	S1 OF REPEAT II (POTENTIAL).
DOMAIN	620	632	EXTRACELLULAR (POTENTIAL).
TRANSMEM	633	654	S2 OF REPEAT II (POTENTIAL).
DOMAIN	655	660	CYTOPLASMIC (POTENTIAL).
TRANSMEM	661	679	S3 OF REPEAT II (POTENTIAL).
DOMAIN	680	687	EXTRACELLULAR (POTENTIAL).
TRANSMEM	688	711	S4 OF REPEAT II (POTENTIAL).
DOMAIN	712	722	CYTOPLASMIC (POTENTIAL).
TRANSMEM	723	743	S5 OF REPEAT II (POTENTIAL).
DOMAIN	744	795	EXTRACELLULAR (POTENTIAL).
TRANSMEM	796	820	S6 OF REPEAT II (POTENTIAL).
DOMAIN	821	1125	CYTOPLASMIC (POTENTIAL).
TRANSMEM	1126	1148	S1 OF REPEAT III (POTENTIAL).
DOMAIN	1149	1166	EXTRACELLULAR (POTENTIAL).
TRANSMEM	1167	1187	S2 OF REPEAT III (POTENTIAL).
DOMAIN	1188	1197	CYTOPLASMIC (POTENTIAL).
TRANSMEM	1198	1217	S3 OF REPEAT III (POTENTIAL).
DOMAIN	1218	1231	EXTRACELLULAR (POTENTIAL).
TRANSMEM	1232	1253	S4 OF REPEAT III (POTENTIAL).
DOMAIN	1254	1263	CYTOPLASMIC (POTENTIAL).
TRANSMEM	1264	1287	S5 OF REPEAT III (POTENTIAL).
DOMAIN	1288	1364	EXTRACELLULAR (POTENTIAL).
TRANSMEM	1365	1390	S6 OF REPEAT III (POTENTIAL).
DOMAIN	1391	1445	CYTOPLASMIC (POTENTIAL).
TRANSMEM	1446	1466	S1 OF REPEAT IV (POTENTIAL).
DOMAIN	1467	1480	EXTRACELLULAR (POTENTIAL).
TRANSMEM	1481	1502	S2 OF REPEAT IV (POTENTIAL).
DOMAIN	1503	1509	CYTOPLASMIC (POTENTIAL).
TRANSMEM	1510	1528	S3 OF REPEAT IV (POTENTIAL).
DOMAIN	1529	1542	EXTRACELLULAR (POTENTIAL).
TRANSMEM	1543	1566	S4 OF REPEAT IV (POTENTIAL).
DOMAIN	1567	1580	CYTOPLASMIC (POTENTIAL).
TRANSMEM	1581	1601	S5 OF REPEAT IV (POTENTIAL).
DOMAIN	1602	1664	EXTRACELLULAR (POTENTIAL).
TRANSMEM	1665	1692	S6 OF REPEAT IV (POTENTIAL).
DOMAIN	1693	1835	CYTOPLASMIC (POTENTIAL).
SITE	355	555	CALCIUM ION SELECTIVITY AND PERMEABILITY (BY SIMILARITY).
SITE	779	779	CALCIUM ION SELECTIVITY AND PERMEABILITY (BY SIMILARITY).
SITE	1339	1339	CALCIUM ION SELECTIVITY AND PERMEABILITY (BY SIMILARITY).
SITE	1637	1637	CALCIUM ION SELECTIVITY AND PERMEABILITY (BY SIMILARITY).
CARBOHYD	171	171	N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD	242	242	N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD	309	309	N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD	1301	1301	N-LINKED (GLCNAC. . .) (POTENTIAL).
CARBOHYD	1304	1304	N-LINKED (GLCNAC. . .) (POTENTIAL).
CONFLICT	193	193	M -> L (IN REF. 2).
CONFLICT	291	291	G -> V (IN REF. 2).
CONFLICT	345	345	G -> W (IN REF. 2).
CONFLICT	394	394	V -> L (IN REF. 2).
CONFLICT	405	405	E -> N (IN REF. 2).
CONFLICT	485	485	C -> S (IN REF. 2).
CONFLICT	512	512	D -> Y (IN REF. 2).
CONFLICT	558	558	S -> R (IN REF. 2).
CONFLICT	683	683	P -> S (IN REF. 2).
CONFLICT	691	691	F -> S (IN REF. 2).
CONFLICT	739	740	MH -> ID (IN REF. 2).
CONFLICT	833	833	C -> Y (IN REF. 2).
CONFLICT	846	846	F -> L (IN REF. 2).
CONFLICT	856	856	S -> R (IN REF. 2).
CONFLICT	905	905	L -> R (IN REF. 2).
CONFLICT	913	913	M -> YW (IN REF. 2).

FT CONFLICT 936 936 W -> G (IN REF. 2).  
FT CONFLICT 996 996 A -> R (IN REF. 2).  
FT CONFLICT 1060 1060 I -> M (IN REF. 2).  
FT CONFLICT 1094 1094 D -> CC (IN REF. 2).  
FT CONFLICT 1197 1198 SS -> TD (IN REF. 2).  
FT CONFLICT 1230 1232 MISSING (IN REF. 2).  
FT CONFLICT 1422 1422 K -> Y (IN REF. 2).  
FT CONFLICT 1623 1625 FGM -> SAR (IN REF. 2).  
SQ SEQUENCE 1835 AA; 205207 MW; 850807A68B82COBE CRC64;  
  
Query Match 78.6%; Score 33; DB 1; Length 1835;  
Best Local Similarity 83.3%; Pred. No. 1.2e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 WRQTRK 6  
DB 584 WRQTRK 589  
  
RESULT 5  
UL38 HCMVA  
ID UL38 HCMVA STANDARD; PRT; 331 AA.  
AC P16779;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-AUG-1990 (Rel. 15, Last sequence update)  
DT 01-FEB-1991 (Rel. 17, Last annotation update)  
DE Hypothetical protein UL38.  
GN UL38.  
OS Human cytomegalovirus (strain AD169).  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Betaherpesvirinae; Cytomegalovirus.  
OX NCBI\_TaxID=10360;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90269039; PubMed=2161319;  
RA Chee M.S., Bankier A.T., Beck S., Bohni R., Brown C.M., Cerny R.,  
Horanell T., Hutchison C.A. III, Kouzarides T., Martignetti J.A.,  
Predie E., Satchwell S.C., Tomlinson P., Weston K.M., Barrell B.G.,  
"Analysis of the protein-coding content of the sequence of human  
cytomegalovirus strain AD169.";  
RT Curr. Top. Microbiol. Immunol. 154:125-169(1990).  
RL  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; X17403; CAA35397.1; -.  
DR PIR; S09800; S09800.  
KW Hypothetical protein.  
FT CARBOHYD 76 76 N-LINKED (GLCNAC... ) (POTENTIAL).  
SQ SEQUENCE 331 AA; 36737 MW; BOAFAD6DA2FF2B49 CRC64;  
  
Query Match 76.2%; Score 32; DB 1; Length 331;  
Best Local Similarity 71.4%; Pred. No. 30;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 WRQTRK 7  
DB 20 WRQTRK 26  
  
RESULT 6  
CBPA ANOGA  
ID CBPA ANOGA STANDARD; PRT; 433 AA.  
AC O02350;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Zinc carboxypeptidase A precursor (EC 3.4.17.-).

OS Anopheles gambiae (African malaria mosquito).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.  
OX NCBI\_TaxID=7185;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=G3;  
RA MEDLINE=982311106; PubMed=9569647;  
RA Edwards M.J., Lemos F.J., Donnelly-Doman M., Jacobs-Lorena M.;  
RT "Rapid induction by a blood meal of a carboxypeptidase gene in the  
gut of the mosquito Anopheles gambiae.";  
RL Insect Biochem. Mol. Biol. 27:1063-1072(1997).  
CC -!- FUNCTION: Involved in the digestion of the blood meal (Probable).  
CC -!- SUBCELLULAR LOCATION: Secreted.  
CC -!- TISSUE SPECIFICITY: Midgut-specific.  
CC -!- INDUCTION: By blood meal.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; AF000953; AAB96576.1; -.  
DR HSSP; P00730; 1PYT.  
DR MEROPS; M14.0PA; -.  
DR InterPro; IPR003146; Propep M14.  
DR InterPro; IPR000834; Zn\_carbOpept.  
DR Pfam; PF02244; Propep\_M14; 1.  
DR Pfam; PF00246; Zn\_carbOpept; 1.  
DR PRINTS; PR00765; CRBOXYPTASEA.  
DR SMART; SM00631; Zn\_pept; 1.  
DR PROSITE; PS00132; CARBOXYPEPT\_ZN\_1; 1.  
DR PROSITE; PS00133; CARBOXYPEPT\_ZN\_2; 1.  
KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Signal.  
FT SIGNAL 1 28  
FT CHAIN 29 433 ZINC CARBOXYPEPTIDASE A.  
FT METAL 187 187 ZINC (BY SIMILARITY).  
FT METAL 190 190 ZINC (BY SIMILARITY).  
FT METAL 312 312 ZINC (BY SIMILARITY).  
FT ACT\_SITE 364 364 PROTON DONOR (BY SIMILARITY).  
FT ACT\_SITE 387 387 NUCLEOPHILE (BY SIMILARITY).  
FT DISULFID 253 276 BY SIMILARITY.  
SQ SEQUENCE 433 AA; 48972 MW; 89D00BEC0DC0DF2 CRC64;  
  
Query Match 76.2%; Score 32; DB 1; Length 433;  
Best Local Similarity 83.3%; Pred. No. 40;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 WRQTRK 6  
DB 242 WRQTRK 247  
  
RESULT 7  
EF11 CRIGR  
ID EF11 CRIGR STANDARD; PRT; 462 AA.  
AC P20001;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Elongation factor 1-alpha 1 (EF-1-alpha-1) (Elongation factor 1 A-1)  
DE (EF1A-1) (Elongation factor Tu) (EF-Tu).  
GN EF1A1 OR EF1A.  
OS Cricetus griseus (Chinese hamster), and  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
OC Cricetulus.  
OX NCBI\_TaxID=10029, 10116;

[1]  
RP SEQUENCE FROM N.A.  
RC SPECIES=C.griseus; STRAIN=V79-UP;  
RX MEDLINE=90110048; PubMed=2606910;  
RA Hayashi Y., Urae R., Utsumi S., Kito M.;  
RT "Anchoring of peptide elongation factor EF-1 alpha by  
RT phosphatidylinositol at the endoplasmic reticulum membrane.";  
J. Biochem. 106:560-563 (1989).  
[2]  
RP SEQUENCE FROM N.A.  
RC SPECIES=HUMAN;  
RX MEDLINE=92178991; PubMed=1542580;  
RA Shiraawa T., Sakamoto K., Akashi T., Takahashi H., Kawashima A.;  
RT "Nucleotide sequence of rat elongation factor-1 alpha cDNA.";  
Nucleic Acids Res. 20:909-909 (1992).  
[3]  
RP SEQUENCE FROM N.A.  
RC SPECIES=HUMAN;  
RX MEDLINE=92107686; PubMed=1762922;  
RA Taniguchi S., Miyamoto S., Hirayuki S., Kobayashi H.;  
RT "Rat elongation factor 1 alpha: sequence of cDNA from a highly  
RT metastatic fos-transferred cell line.";  
Nucleic Acids Res. 19:6949-6949 (1991).  
CC -!- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT BINDING OF  
CC AMINOACYL-TRNA TO THE A-SITE OF RIBOSOMES DURING PROTEIN  
CC BIOSYNTHESIS.  
CC -!- SUBCELLULAR LOCATION: ANCHORED AT THE ENDOPLASMIC RETICULUM  
CC MEMBRANE BY PHOSPHATIDYLINOSITOL VIA ETHANOLAMINE BRIDGING.  
CC -!- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.  
CC EF-TU/EF-1A SUBFAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; D00522; BA000409.1; -;  
DR EMBL; X63561; CA445122.1; -;  
DR EMBL; X61043; CA443378.1; -;  
DR PIR; J00133; J00133.  
DR PIR; S21055; S21055.  
DR HSP; P07157; 1A1P.  
DR InterPro; IPR004539; EF1\_alpha.  
DR InterPro; IPR000795; EF\_GTPbind.  
DR InterPro; IPR004160; EFTU\_Cterm.  
DR InterPro; IPR004161; EFTU\_D2.  
DR Pfam; PF00009; GTP\_EFTU; 1.  
DR Pfam; PF03144; GTP\_EFTU\_D2; 1.  
DR Pfam; PF03143; GTP\_EFTU\_D3; 1.  
DR PRINTS; PR00315; ELONGATNFCF.  
DR TIGRfam; TIGR00483; EF-1\_alpha; 1.  
DR PROSITE; PS00301; EFACITOR\_GTP; 1.  
KW Elongation factor; Protein biosynthesis; GTP-binding; Methylation.  
FT NP\_BIND 14 21  
FT GTP (BY SIMILARITY).  
FT NP\_BIND 91 95  
FT GTP (BY SIMILARITY).  
FT NP\_BIND 153 156  
FT GTP (BY SIMILARITY).  
FT MOD\_RES 36 36 METHYLATION (TRI-) (BY SIMILARITY).  
FT MOD\_RES 55 55 METHYLATION (DI-) (BY SIMILARITY).  
FT MOD\_RES 79 79 METHYLATION (TRI-) (BY SIMILARITY).  
FT MOD\_RES 165 165 METHYLATION (DI-) (BY SIMILARITY).  
FT MOD\_RES 318 318 METHYLATION (TRI-) (BY SIMILARITY).  
FT BINDING 301 301 ETHANOLAMINE-PHOSPHOGLYCEROL.  
FT BINDING 374 374 ETHANOLAMINE-PHOSPHOGLYCEROL.  
SQ SEQUENCE 462 AA; 50114 MW; 71072871DE7405DC CRC64;  
Query Match 76.2%; Score 32; DB 1; Length 462;  
Best Local Similarity 71.4%; Pred.No. 43;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
1 WRQTRKD 7

Db 214 WKVTRKD 220  
: |||||  
RESULT 8  
EF1L HUMAN  
ID EF1L HUMAN STANDARD; PRT; 462 AA.  
AC P04720; P04719;  
DT 13-AUG-1987 (Rel. 05, Created)  
DT 13-AUG-1987 (Rel. 05, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Elongation factor 1-alpha 1 (EF-1-alpha-1) (Elongation factor 1 A-1)  
DE (EF1A-1) (Elongation factor Tu) (EF-Tu).  
GN EF1A1 OR EF1A OR EF1A.  
OS Homo sapiens (Human),  
OS Bos taurus (Bovine), and  
OS Oryctolagus cuniculus (Rabbit).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606, 9913, 9986;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES=Human;  
RX MEDLINE=86136120; PubMed=3512269;  
RA Brands J.H.G.M., Maassen J.A., van Hemert F.J., Amons R., Moeller W.;  
RT "The primary structure of the alpha subunit of human elongation  
RT factor 1. Structural aspects of guanine-nucleotide-binding sites.";  
Eur. J. Biochem. 155:167-171 (1986).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC SPECIES=Human;  
RX MEDLINE=89174636; PubMed=2564392;  
RA Uetsuki T., Naito A., Nagata S., Kaziro Y.;  
RT "Isolation and characterization of the human chromosomal gene for  
RT polypeptide chain elongation factor-1 alpha.";  
J. Biol. Chem. 264:5791-5798 (1989).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC SPECIES=Human; TISSUE=Liver;  
RX MEDLINE=90221877; PubMed=2183196;  
RA Madsen H.O., Poulsen K., Dahl O., Clark B.F.C., Hjorth J.P.;  
RT "Retropseudogenes constitute the major part of the human elongation  
RT factor 1 alpha gene family.";  
Nucleic Acids Res. 18:1513-1516 (1990).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC SPECIES=Human;  
RX Shimazu T., Koike K.;  
RT "Postnatal expression of a novel mRNA isoform from the human  
RT elongation factor-1a gene.";  
Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A.  
RC SPECIES=Human;  
RX Tissue=B-cell, Colon, Lung, Lymph, Muscle, Ovary, Pancreas,  
RC Placenta, and Uterus;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Sapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaney S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Vallalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko V., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,

RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RL human and mouse cDNA sequences.";  
 RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 RP [6]  
 RC SEQUENCE OF 1-94 FROM N.A.  
 RX SPECIES=Human;  
 RA MEDLINE=86176739; PubMed=3960725;  
 RL Rao T.R., Slobin L.I.;  
 RT "Structure of the amino-terminal end of mammalian elongation factor  
 RL Tu.";  
 RN Nucleic Acids Res. 14:2409-2409 (1986).  
 RP [7]  
 RC SEQUENCE OF 138-462 FROM N.A.  
 RX SPECIES=Human;  
 RA MEDLINE=88153640; PubMed=3346208;  
 RL Ann D.K., Wu M.M.J., Huang T., Carlson D.M., Wu R.;  
 RT "Retinol-regulated gene expression in human tracheobronchial  
 RL epithelial cells. Enhanced expression of elongation factor EF-1  
 RL alpha.";  
 RN J. Biol. Chem. 263:3546-3549 (1988).  
 RP [8]  
 RC SEQUENCE FROM N.A.  
 RX SPECIES=Bovine;  
 RA Perez D.R., Johnson C.M., Donis R.O.;  
 RT "BVDV NS5A interacts with eEF1a.";  
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
 RN [9]  
 RC SEQUENCE FROM N.A.  
 RX SPECIES=Bovine; STRAIN=Japanese black; TISSUE=Endometrium;  
 RA Kojima T., Oshima K., Watanabe H., Komatsu M.;  
 RT "cDNA of bovine elongation factor 1 alpha from endometrium.";  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 RN [10]  
 RC SEQUENCE FROM N.A.  
 RX SPECIES=Rabbit; TISSUE=Liver;  
 RA MEDLINE=92220623; PubMed=1561101;  
 RL Cavallius J., Merrick W.C.;  
 RT "Nucleotide sequence of rabbit elongation factor 1 alpha cDNA.";  
 RN Nucleic Acids Res. 20:1422-1422 (1992).  
 RP [11]  
 RC SEQUENCE FROM N.A.  
 RX SPECIES=Rabbit; STRAIN=New Zealand white; TISSUE=Spleen;  
 RA Sheu G., Traugh J.A.;  
 RT Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.  
 RN [12]  
 RC PARTIAL SEQUENCE, AND POST-TRANSLATIONAL MODIFICATIONS.  
 RX SPECIES=Rabbit;  
 RA MEDLINE=90062188; PubMed=2511205;  
 RL Dever T.E., Costello C.B., Owens C.L., Rosenberry T.L., Merrick W.C.;  
 RT "Location of seven post-translational modifications in rabbit  
 RL elongation factor 1 alpha including dimethyllysine, trimethyllysine,  
 RL and glycerylphosphorylethanolamine.";  
 RN J. Biol. Chem. 264:20518-20525 (1989).  
 RP [13]  
 RC ETHANOLAMINE BINDING SITES.  
 RX SPECIES=Human;  
 RA MEDLINE=89340549; PubMed=2569467;  
 RL Whiteheart S.W., Shenbagarmurthi P., Chen L., Cotter R.J.,  
 RA Hart G.W.;  
 RT "Murine elongation factor 1 alpha (EF-1 alpha) is posttranslationally  
 RT modified by novel amide-linked ethanolamine-phosphoglycerol moieties.  
 RT Addition of ethanolamine-phosphoglycerol to specific glutamic acid  
 RT residues on EF-1 alpha.";  
 RN J. Biol. Chem. 264:14334-14341 (1989).  
 CC -!- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT BINDING OF  
 CC AMINOACYL-TRNA TO THE A-SITE OF RIBOSOMES DURING PROTEIN  
 CC BIOSYNTHESIS.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -!- TISSUE SPECIFICITY: BRAIN, PLACENTA, LUNG, LIVER, KIDNEY,  
 CC PANCREAS BUT BARELY DETECTABLE IN HEART AND SKELETAL MUSCLE.  
 CC -!- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.  
 CC EF-TU/EF-1A SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
 CC -----  
 DR EMBL; X03558; CAA27245.1; -  
 DR EMBL; X03689; CAA27325.1; -  
 DR EMBL; X16869; CAA34756.1; -  
 DR EMBL; X043301; RAK95378.1; -  
 DR EMBL; BC02845; AAH02845.1; -  
 DR EMBL; BC006102; AAH06102.1; -  
 DR EMBL; BC008587; AAH08587.1; -  
 DR EMBL; BC009733; AAH09733.1; -  
 DR EMBL; BC009875; AAH09875.1; -  
 DR EMBL; BC010735; AAH10735.1; -  
 DR EMBL; BC012891; AAH12891.1; -  
 DR EMBL; BC014224; AAH14224.1; -  
 DR EMBL; BC018150; AAH18150.1; -  
 DR EMBL; BC018641; AAH18641.1; -  
 DR EMBL; BC021686; AAH21686.1; -  
 DR EMBL; M29548; AAH2367.1; ALT\_INIT.  
 DR EMBL; J04617; AAH52343.1; -  
 DR EMBL; AJ238405; CAB88863.1; -  
 DR EMBL; AS060107; BAB60846.1; -  
 DR EMBL; X62245; CAA44162.1; -  
 DR EMBL; U09823; AAA18502.1; -  
 DR PIR; B24977; EFHUI.  
 DR PIR; S22583; EFRBL.  
 DR HSSP; P07157; IAI.  
 DR SWISS-2DPAGE; P04720; HUMAN.  
 DR Genew; HGNC:3189; EEF1A.  
 DR MIM; 130590; -  
 DR GO; GO:0005853; C:eukaryotic translation elongation factor 1...; TAS.  
 DR GO; GO:0005525; F:GTP binding activity; TAS.  
 DR GO; GO:0006414; P:translational elongation; TAS.  
 DR InterPro; IPR004539; EF1\_alpha.  
 DR InterPro; IPR000795; EF\_GTPbind.  
 DR InterPro; IPR004160; EFTU\_Cterm.  
 DR InterPro; IPR004161; EFTU\_D2.  
 DR Pfam; PF00009; GTP\_EFTU; 1.  
 DR Pfam; PF03144; GTP\_EFTU\_D2; 1.  
 DR Pfam; PF03143; GTP\_EFTU\_D3; 1.  
 DR PRINTS; PR00315; ELONGATNFCT.  
 DR TIGRFS; TIGR00483; EF-1\_alpha; 1.  
 DR PROSITE; PS00301; EFACITOR\_GTP; 1.  
 KW Elongation factor; Protein biosynthesis; GTP-binding; Methylation;  
 KW Multigene family.  
 FT NP\_BIND 14 21 GTP (BY SIMILARITY).  
 FT NP\_BIND 91 95 GTP (BY SIMILARITY).  
 FT NP\_BIND 153 156 GTP (BY SIMILARITY).  
 FT MOD\_RES 36 36 METHYLATION (TRI-).  
 FT MOD\_RES 55 55 METHYLATION (DI-).  
 FT MOD\_RES 79 79 METHYLATION (DI-).  
 FT MOD\_RES 165 165 METHYLATION (DI-).  
 FT MOD\_RES 318 318 METHYLATION (TRI-).  
 FT BINDING 301 301 ETHANOLAMINE-PHOSPHOGLYCEROL.  
 Query Match 76.2%; Score 32; DB 1; Length 462;  
 Best Local Similarity 71.4%; Pred. No. 43;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 WRQTRKD 7  
 DB 214 WKVTRKD 220  
 RESULT 9  
 EFTU\_MOUSE  
 ID EFTU\_MOUSE STANDARD; PRT; 462 AA.

AC P10126; O61511;  
 DT 01-MAR-1989 (Rel. 10, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Elongation factor 1-alpha 1 (EF-1-alpha-1) (Elongation factor 1 A-1)  
 DE (EF1A-1) (Elongation factor Tu) (EF-Tu).  
 GN EF1A1 OR EF1A.  
 OS Mus musculus (Mouse).  
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RP MEDLINE=89108007; PubMed=3215517;  
 RA Reddy N.S., Roth W.W., Bragg P.W., Wahba A.J.;  
 RT "Isolation and mapping of a gene for protein synthesis initiation  
 RT factor 4A and its expression during differentiation of murine  
 RT erythroleukemia cells.";  
 RL Gene 70:231-243 (1988).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RP MEDLINE=89098401; PubMed=2911475;  
 RA Lu X., Werner D.;  
 RT "The complete cDNA sequence of mouse elongation factor 1 alpha (EF 1  
 RT alpha) mRNA.";  
 RL Nucleic Acids Res. 17:442-442 (1989).  
 RN [3]  
 RP SEQUENCE OF 1-94 FROM N.A.  
 RP MEDLINE=86176739; PubMed=3960725;  
 RA Rao T.R., Slobin L.I.;  
 RT "Structure of the amino-terminal end of mammalian elongation factor  
 RT Tu.";  
 RL Nucleic Acids Res. 14:2409-2409 (1986).  
 RN [4]  
 RP SEQUENCE OF 1-108 FROM N.A.  
 RP MEDLINE=88122115; PubMed=3481036;  
 RA Roth W.W., Bragg P.W., Corrias M.V., Reddy N.S., Dholakia J.N.,  
 RA Wahba A.J.;  
 RT "Expression of a gene for mouse eucaryotic elongation factor Tu during  
 RT murine erythroleukemic cell differentiation.";  
 RL Mol. Cell. Biol. 7:3929-3936 (1987).  
 RN [5]  
 RP ETHANOLAMINE BINDING SITES, AND SEQUENCE OF 291-313 AND 372-376.  
 RP MEDLINE=89340549; PubMed=2569467;  
 RA Whiteheart S.W., Shenbagamurthi P., Chen L., Cotter R.J.,  
 RA White G.W.;  
 RT "Murine elongation factor 1 alpha (EF-1 alpha) is posttranslationally  
 RT modified by novel amide-linked ethanolamine-phosphoglycerol moieties.  
 RT Addition of ethanolamine-phosphoglycerol to specific glutamic acid  
 RT residues on EF-1 alpha.";  
 RL J. Biol. Chem. 264:14334-14341 (1989).  
 CC -!- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT BINDING OF  
 CC AMINOACYL-TRNA TO THE A-SITE OF RIBOSOMES DURING PROTEIN  
 CC BIOSYNTHESIS.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -!- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.  
 CC EF-TU/EF-1A SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M22432; AAA50406.1; -  
 CC EMBL; X13661; CAA31957.1; -  
 CC EMBL; X03688; CAA27324.1; -  
 CC EMBL; M17878; AAA37538.1; -  
 CC F01; S02114; EFW51.  
 CC PIR; P07157; IAIIP.  
 CC HSP; P07157; IAIIP.  
 CC SWISS-2DPAGE; P10126; MOUSE.

DR MGD; MGI:1096881; Ref1a1.  
 DR InterPro; IPR004539; EFl\_alpha.  
 DR InterPro; IPR000795; EF\_GTPbind.  
 DR InterPro; IPR004160; EFTU\_Cterm.  
 DR InterPro; IPR004161; EFTU\_D2.  
 DR Pfam; PF000039; GTP\_EFTU\_1.  
 DR Pfam; PF03144; GTP\_EFTU\_D2; 1.  
 DR Pfam; PF03143; GTP\_EFTU\_D3; 1.  
 DR PRINTS; PR00315; ELONGATNFTC.  
 DR TIGRFAMs; TIGR00483; EF-1\_alpha; 1.  
 DR PROSITE; PS00301; EFACTOR\_GTP; 1.  
 KW Elongation factor; Protein biosynthesis; GTP-binding; Methylation;  
 KW Multigene family.  
 FT NP\_BIND 14 21 GTP (BY SIMILARITY).  
 FT NP\_BIND 91 95 GTP (BY SIMILARITY).  
 FT NP\_BIND 153 156 GTP (BY SIMILARITY).  
 FT MOD\_RES 36 36 METHYLATION (TRI-) (BY SIMILARITY).  
 FT MOD\_RES 55 55 METHYLATION (DI-) (BY SIMILARITY).  
 FT MOD\_RES 79 79 METHYLATION (DI-) (BY SIMILARITY).  
 FT MOD\_RES 165 165 METHYLATION (DI-) (BY SIMILARITY).  
 FT MOD\_RES 318 318 ETHANOLAMINE-PHOSPHOGLYCEROL.  
 FT BINDING 301 301 ETHANOLAMINE-PHOSPHOGLYCEROL.  
 FT BINDING 374 374 H -> R (IN REF. 4).  
 FT CONFLICT 7 7 H -> L (IN REF. 4).  
 FT CONFLICT 15 15 T -> S (IN REF. 4).  
 FT CONFLICT 23 23 LW -> QR (IN REF. 2).  
 FT CONFLICT 77 78 S -> A (IN REF. 3).  
 FT CONFLICT 83 83 DA -> ES (IN REF. 2).  
 FT CONFLICT 91 92 O -> R (IN REF. 4).  
 FT CONFLICT 108 108 HASGT -> SAVAP (IN REF. 2).  
 FT CONFLICT 222 226 MISSING (IN REF. 2).  
 FT CONFLICT 239 239  
 SQ SEQUENCE 462 AA; 50164 MW; 75572871D57405DC CRC64;  
 Query Match 76.2%; Score 32; DB 1; Length 462;  
 Best Local Similarity 71.4%; Pred. No. 43;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 WRQTRKD 7  
 Db 214 WKVTRKD 220  
 RESULT 10  
 EF1A\_CHICK STANDARD; PRT; 462 AA.  
 ID EF1A\_CHICK  
 AC Q90835;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Elongation factor 1-alpha 1 (EF-1-alpha-1) (Elongation factor Tu) (EF-Tu).  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RP STRAIN=White leghorn; TISSUE=Liver;  
 RX MEDLINE=94192993; PubMed=8144022;  
 RA Wang H., Parent M., Morais R.;  
 RT "Cloning and characterization of a cDNA encoding elongation factor 1  
 RT alpha from chicken cells devoid of mitochondrial DNA.";  
 RL Gene 140:155-161 (1994).  
 CC -!- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT BINDING OF  
 CC AMINOACYL-TRNA TO THE A-SITE OF RIBOSOMES DURING PROTEIN  
 CC BIOSYNTHESIS.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -!- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.  
 CC EF-TU/EF-1A SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

DR EMBL: L00677; AAA48757.1; --  
DR PIR: I50226; I50226.  
DR HSP: P07157; IAP.  
DR InterPro: IPR004539; EFL\_alpha.  
DR InterPro: IPR000795; EF\_GTPbind.  
DR InterPro: IPR004160; EFTU\_Cterm.  
DR InterPro: IPR004161; EFTU\_D2.  
DR Pfam: PF00009; GTP\_EFTU; 1.  
DR Pfam: PF03144; GTP\_EFTU\_D2; 1.  
DR Pfam: PF03143; GTP\_EFTU\_D3; 1.  
DR PRINTS: PR00315; ELONGATNFT.  
DR TIGRFS: TIGR00483; EF-1\_alpha; 1.  
DR PROSITE: PS00301; EFATOR\_GTP; 1.  
DR Elongation factor; Protein biosynthesis; GTP-binding.  
FT NP\_BIND 14 21 GTP (BY SIMILARITY).  
FT NP\_BIND 91 95 GTP (BY SIMILARITY).  
FT NP\_BIND 153 156 GTP (BY SIMILARITY).  
SQ SEQUENCE 462 AA; 50157 MW; D6C0BE55540A686A CRC64;

Query Match 76.2%; Score 32; DB 1; Length 462;  
Best Local Similarity 71.4%; Pred. No. 43;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 WRQTRKD 7  
|:||||  
DB 214 WKVTRKD 220

#### RESULT 11

ID C7C3 MAIZE STANDARD; PRT; 535 AA.  
AC P93703; Q43256;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Cytochrome P450 71C3 (EC 1.14.-.-).  
GN CYP71C3  
OS Zea mays (Maize).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC PACCAD clade; Panicoideae; Andropogoneae; Zea.  
OX NCBI\_TaxID=4577;  
RN [1]  
RP SEQUENCE OF 2-535 FROM N.A.  
RC STRAIN=CV\_C131A;  
RX MEDLINE=95124291; PubMed=7823905;  
RA Frey M., Klem R., Saedler H., Gierl A.;  
RT "Expression of a cytochrome P450 gene family in maize."  
RL Mol. Gen. Genet. 246:100-109(1995).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV\_C131A;  
RA Gierl A.;  
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: Belongs to the cytochrome P450 family.  
-----

This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

DR EMBL: X81830; CAA57424.2; ALT\_SEQ.  
DR EMBL: Y11403; CAA72207.1; --

DR PIR: T03246; T03246.  
DR InterPro: IPR001128; Cytochrome\_P450.  
DR Pfam: PF00067; P450; 1.  
DR PRINTS: PR00385; P450.  
DR PROSITE: PS00086; CYTOCHROME\_P450; 1.  
KW Oxidoreductase; Monooxygenase; Transmembrane; Heme.  
FT TRANSMEM 23 43 POTENTIAL.  
FT METAL 475 475 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).  
SQ SEQUENCE 535 AA; 60715 MW; F92A696108E2ADAF CRC64;

Query Match 76.2%; Score 32; DB 1; Length 535;  
Best Local Similarity 83.3%; Pred. No. 51;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WRQTRK 6  
|:||||  
DB 151 WQTRK 156

#### RESULT 12

ID AL11 ARAHY STANDARD; PRT; 614 AA.  
AC P43237;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Allergen Ara h 1, Clone P17 precursor (Ara h 1).  
OS Arachis hypogaea (peanut).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids I; Fabaales; Fabaceae; Papilionoideae; Aeschynomeneae;  
OC Arachis.  
OX NCBI\_TaxID=3818;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV\_Florum;  
RX MEDLINE=96013631; PubMed=7560062;  
RA Burks A.W., Cockrell G., Stanley J.S., Helm R.M., Bannon G.A.;  
RT "Recombinant peanut allergen Ara h 1 expression and IgE binding in  
RT patients with peanut hypersensitivity."  
RL J. Clin. Invest. 96:1715-1721(1995).  
RN [2]  
RP CARBOHYDRATE-LINKAGE SITE ASN-516.  
RX MEDLINE=20455243; PubMed=10998264;  
RA Kolarich D., Altmann F.;  
RT "N-glycan analysis by matrix-assisted laser desorption/ionization mass  
RT spectrometry of electrophoretically separated nonmammalian proteins:  
RT application to peanut allergen Ara h 1 and olive pollen allergen Ole  
RT e 1."  
RL Anal. Biochem. 285:64-75(2000).  
CC -!- SIMILARITY: Belongs to the 7S seed storage protein family.  
-----

This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

EMBL: L38853; AAA60336.1; --  
DR HSP: P50477; 1CAU.  
DR GlycoSuiteDB; P43237; --  
DR InterPro: IPR006045; Cupin.  
DR InterPro: IPR007113; Cupin\_sup.  
DR Pfam: PF00190; Cupin; 2.  
KW Allergen; Glycoprotein; Signal.  
FT SIGNAL 1 25 POTENTIAL.  
FT CHAIN 26 614 ALLERGEN ARA H 1, CLONE P17.  
FT CARBOHYD 516 516 N-LINKED (GLCNAC...)  
SQ SEQUENCE 614 AA; 70283 MW; 1DDACF217ECSF31 CRC64;

Query Match 76.2%; Score 32; DB 1; Length 614;  
Best Local Similarity 71.4%; Pred. No. 59;  
Matches 5; Conservative 1; Mismatches 0; Gaps 0;  
Oy 1 WRQTRKD 7  
Db 123 WRQPRD 129

RESULT 13  
ALL2\_ARAHY STANDARD; PRT; 626 AA.  
AC P43238;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
D3 Allergen Ara h 1, clone P41B precursor (Ara h I).  
OS Arachis hypogaea (Peanut).  
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eucosids I; Fabales; Fabaceae; Papilionoideae; Aeschynomeneae;  
OC Arachis.  
OX NCBI\_TaxID=3818;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Florunner;  
RX MEDLINE=96013631; PubMed=7560062;  
RA Burks A.W., Cockrell G., Stanley J.S., Helm R.M., Bannon G.A.;  
RT "Recombinant peanut allergen Ara h I expression and IgE binding in  
RT patients with peanut hypersensitivity.";  
RL J. Clin. Invest. 96:1715-1721(1995).  
RN [2]  
RP CARBOHYDRATE-LINKAGE SITE ASN-521.  
RX MEDLINE=20455243; PubMed=1099264;  
RA Kolarich D., Altmann F.;  
RT "N-glycan analysis by matrix-assisted laser desorption/ionization mass  
RT spectrometry of electrophoretically separated nonmammalian proteins:  
RT application to peanut allergen Ara h I and olive pollen allergen Ole  
RT e 1.";  
RL Anal. Biochem. 285:64-75(2000).  
CC -1- SIMILARITY: belongs to the 7S seed storage protein family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; L34402; AAC00861.1; -  
DR HSSP; P50477; 1CAU.  
DR InterPro; IPR006045; Cupin.  
DR InterPro; IPR007113; Cupin\_sup.  
DR Pfam; PF00190; Cupin; 2.  
KW Allergen; Glycoprotein; Signal.  
FT SIGNAL 1 25 POTENTIAL.  
FT CHAIN 26 626 ALLERGEN ARA H 1, CLONE P41B.  
FT CARBOHYD 521 521 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 626 AA; 71345 MW; 1A6BBB41490D0E3 CRC64;  
Query Match 76.2%; Score 32; DB 1; Length 626;  
Best Local Similarity 71.4%; Pred. No. 60;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
Oy 1 WRQTRKD 7  
Db 129 WRQPRD 135

RESULT 14  
TRG1\_ECOLI STANDARD; PRT; 938 AA.  
ID TRG1\_ECOLI

P33790;  
01-FEB-1994 (Rel. 28, Created)  
01-FEB-1994 (Rel. 28, Last sequence update)  
16-OCT-2001 (Rel. 40, Last annotation update)  
DT Trag protein.  
DE Trag.  
CN Trag.  
OS Escherichia coli.  
OG Plasmid F.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=92204127; PubMed=1348105;  
RA Firth N., Skurray R.A.;  
RT "Characterization of the F plasmid bifunctional conjugation gene,  
RT trag.";  
RL Mol. Gen. Genet. 232:145-153(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94359430; PubMed=7915817;  
RA Frost L.S., Ippen-Ihler K., Skurray R.A.;  
RT "Analysis of the sequence and gene products of the transfer region of  
RT the F sex factor.";  
RL Microbiol. Rev. 58:162-210(1994).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / CR63;  
RA Shimizu H., Saitoh Y., Suda Y., Uehara K., Sampei G., Mizobuchi K.;  
RT "Complete nucleotide sequence of the F plasmid: its implications for  
RT organization and diversification of plasmid genomes.";  
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: PLAYS A CRUCIAL ROLE IN DONOR-RECIPIENT CELL  
CC INTERACTIONS. REQUIRED FOR TWO STAGES OF THE CONJUGATION PROCESS:  
CC PILUS BIOSYNTHESIS AND MATING AGGREGATE STABILIZATION. MAY  
CC INTERACT WITH TRAN.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.  
CC -1- PTM: TRAG\* PROBABLY ARISES FROM THE POST-TRANSLATIONAL CLEAVAGE OF  
CC TRAG.  
CC -1- CAUTION: TRAG IS NOT RESPONSIBLE FOR THE N-TERMINAL ACETYLATION  
CC OF F PILIN AS STATED BY SOME AUTHORS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M59763; AAA98081.1; -  
DR EMBL; U01159; AAC44184.1; -  
DR EMBL; AP001918; BAA97969.1; -  
DR PIR; S20480; S20480.  
DR EcoGene; EG40114; trag.  
KW Complete proteome; Conjugation; Transmembrane; Inner membrane;  
FT CHAIN 1 938 TRAG PROTEIN.  
FT CHAIN ? 938 TRAG\* PROTEIN.  
FT DOMAIN 1 53 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 54 73 POTENTIAL.  
FT DOMAIN 74 329 PERIPLASMIC (POTENTIAL).  
FT TRANSMEM 330 350 POTENTIAL.  
FT DOMAIN 351 412 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 413 433 POTENTIAL.  
FT DOMAIN 434 938 PERIPLASMIC (POTENTIAL).  
SQ SEQUENCE 938 AA; 102471 MW; 9E09C8402ACDFDDF CRC64;  
Query Match 76.2%; Score 32; DB 1; Length 938;  
Best Local Similarity 57.1%; Pred. No. 93;  
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;



QY 1 WROTRKD 7  
 Db 810 WRESRRD 816

RESULT 15  
 PTPJ HUMAN STANDARD; PRT; 1337 AA.  
 AC Q12913; Q15255;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Protein-tyrosine phosphatase eta precursor (BC 3.1.3.48) (R-PTP-eta)  
 DE (HPTP eta) (Density enhanced phosphatase-1) (DEP-1) (CD148 antigen).  
 GN PTPRJ OR DEP1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95024024; PubMed=7937872;  
 RA Oestman A., Yang Q., Tonks N.K.;  
 RT "Expression of DEP-1, a receptor-like protein-tyrosine-phosphatase,  
 is enhanced with increasing cell density.";  
 FT Proc. Natl. Acad. Sci. U.S.A. 91:9680-9684(1994).  
 RL [2]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE=95086212; PubMed=7994032;  
 RA Honda H., Inazawa J., Nishida J., Yazaki Y., Hirai H.;  
 RT "Molecular cloning, characterization, and chromosomal localization of  
 a novel protein-tyrosine phosphatase, HPTP eta.";  
 FT Blood 84:4186-4194(1994).  
 CC -!- FUNCTION: MAY CONTRIBUTE TO THE MECHANISM OF CONTRACT INHIBITION OF  
 CELL GROWTH.  
 CC -!- CATALYTIC ACTIVITY: Protein tyrosine phosphatase + H(2)O = protein  
 tyrosine + phosphate.  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -!- PTM: N- AND O-GLYCOSYLATED.  
 CC -!- SIMILARITY: Contains 1 protein-tyrosine phosphatase domain.  
 CC -!- SIMILARITY: Contains 5 fibronectin type III domains.  
 CC -!- DATABASE: NAME=PRO; NOTE=CD guide CD148 entry;  
 CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd148.htm".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see http://www.isb-sib.ch/announce/  
 or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; U10886; AAB36687.1; -;  
 CC EMBL; D37781; BAA07035.1; -;  
 CC PIR; I38670; I38670.  
 CC HSPG; P18052; IYFO.  
 CC Genew; HGNC:9673; PTPRJ.  
 CC MIM; 600925; -;  
 CC GO; GO:0005887; C:integral to plasma membrane; TAS.  
 CC GO; GO:0005001; P:transmembrane receptor protein tyrosine pho. . .; TAS.  
 CC GO; GO:0007267; P:cell-cell signaling; TAS.  
 CC GO; GO:0006470; P:protein amino acid dephosphorylation; TAS.  
 CC GO; GO:0007169; P:transmembrane receptor protein tyrosine kin. . .; TAS.  
 CC InterPro; IPR003961; FN III.  
 CC InterPro; IPR00387; TVR phosphatase.  
 CC InterPro; IPR00242; Tyr\_PP.  
 CC Pfam; PF00041; fn3; 6.  
 CC Pfam; PF00102; Y\_phosphatase; 1.  
 CC PRINTS; PR00700; PTPRJPHPTASE.  
 CC SMART; SM00060; FN3; 8.  
 CC SMART; SM00194; PTPC; 1.  
 CC PROSITE; PS00383; TVR PHOSPHATASE 1; 1.  
 CC PROSITE; PS00383; TVR PHOSPHATASE 2; 1.  
 CC PROSITE; PS00556; TVR PHOSPHATASE 1; 1.

DR PROSITE; PS00555; TVR PHOSPHATASE PTP; 1.  
 KW Signal; Glycoprotein; Transmembrane; Repeat; Hydrolase.  
 FT SIGNAL 1 35  
 FT CHAIN 36 1337  
 FT DOMAIN 36 975  
 FT TRANSMEM 976 996  
 FT DOMAIN 997 1337  
 FT DOMAIN 119 199  
 FT DOMAIN 366 446  
 FT DOMAIN 454 532  
 FT DOMAIN 540 615  
 FT DOMAIN 626 710  
 FT DOMAIN 1065 1337  
 FT ACT SITE 1239 1337  
 FT CARBOHYD 72 72  
 FT CARBOHYD 82 82  
 FT CARBOHYD 93 93  
 FT CARBOHYD 104 104  
 FT CARBOHYD 142 142  
 FT CARBOHYD 172 172  
 FT CARBOHYD 192 192  
 FT CARBOHYD 231 231  
 FT CARBOHYD 258 258  
 FT CARBOHYD 278 278  
 FT CARBOHYD 342 342  
 FT CARBOHYD 351 351  
 FT CARBOHYD 376 376  
 FT CARBOHYD 391 391  
 FT CARBOHYD 396 396  
 FT CARBOHYD 413 413  
 FT CARBOHYD 431 431  
 FT CARBOHYD 501 501  
 FT CARBOHYD 525 525  
 FT CARBOHYD 536 536  
 FT CARBOHYD 582 582  
 FT CARBOHYD 603 603  
 FT CARBOHYD 618 618  
 FT CARBOHYD 628 628  
 FT CARBOHYD 637 637  
 FT CARBOHYD 666 666  
 FT CARBOHYD 669 669  
 FT CARBOHYD 761 761  
 FT CARBOHYD 772 772  
 FT CARBOHYD 784 784  
 FT CARBOHYD 790 790  
 FT CARBOHYD 824 824  
 FT CARBOHYD 910 910  
 FT CARBOHYD 937 937  
 FT CONFLICT 261 261  
 FT CONFLICT 918 929  
 SQ SEQUENCE 1337 AA; 145985 MW; 5686DE6D1F64236E CRC64;

Query Match 76.2%; Score 32; DB 1; Length 1337;  
 Best Local Similarity 71.4%; Pred. No. 1.4e+02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 WROTRKD 7  
 Db 996 WRESRRD 1002

Search completed: February 11, 2004, 17:04:03  
 Job time : 11.1667 secs

translation elonga  
translation elonga  
translation elonga  
translation elonga  
translation elonga  
cytochrome p450 -  
cytochrome p450 -  
hypothetical prote  
hypothetical prote  
hypothetical prote  
tag protein - bac  
protein-tyrosine-p  
salicylate biosynt  
pchs protein - pse  
probable invertase  
regulatory protein

30 32 76.2 462 1 BPHU1  
31 32 76.2 462 1 EFRB1  
32 32 76.2 462 2 JU0133  
33 32 76.2 462 2 S21053  
34 32 76.2 462 2 I50226  
35 32 76.2 535 2 T03246  
36 32 76.2 535 2 T03260  
37 32 76.2 623 2 T16167  
38 32 76.2 623 2 AD2295  
39 32 76.2 828 2 G87584  
40 32 76.2 938 2 S20480  
41 32 76.2 1337 1 I38670  
42 31 73.8 101 2 A83117  
43 31 73.8 102 2 S58228  
44 31 73.8 146 2 C90673  
45 31 73.8 223 2 S32872

ALIGNMENTS

RESULT 1  
T25542  
hypothetical protein C10H11.10 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C;Accession: T25542  
R;Dante, M.; Wamsley, P.  
submitted to the EMBL Data Library, February 1997  
A;Description: The sequence of C. elegans cosmid C10H11.  
A;Reference number: Z20047  
A;Accession: T25542  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-450 <DAN>  
A;Cross-references: EMBL:U88311; PIDN:ABA42349.1; GSPDB:GN00019; CESP:C10H11.10  
A;Experimental source: strain Bristol N2; clone C10H11  
C;Genetics:  
A;Gene: CESP:C10H11.10  
A;Map position: 1  
A;Introns: 4/2; 218/3; 281/3

Query Match 88.1%; Score 37; DB 2; Length 450;  
Best Local Similarity 85.7%; Pred. No. 8.8;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 WRQTRKD 7  
Db 108 WRQTEKD 114

RESULT 2  
B64184  
oligopeptide transport ATP-binding protein oppF - Haemophilus influenzae (strain Rd KW  
C;Species: Haemophilus influenzae  
C;Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 20-Aug-1999  
C;Accession: B64184  
R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage,  
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman,  
D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.  
Science 285, 496-512, 1995  
A;Authors: Grehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter.  
A;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
A;Reference number: A64000; MUID:95350630; PMID:7542800  
A;Accession: B64184  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-332 <TIGR>  
A;Cross-references: GB:U32792; GB:L42023; NID:G1574672; PIDN:AAC22774.1; PID:G1574675;  
C;Genetics:  
A;Gene: oppF  
C;Function:  
A;Description: probably responsible for energy-coupling to the transport system

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model  
Run on: February 11, 2004, 16:55:29 ; Search time 9.33333 Seconds  
(without alignments)  
72.127 Million cell updates/sec

Title: US-09-901-187C-1  
Perfect score: 42  
Sequence: 1 WRQTRKD 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_76.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	37	88.1	450	2 T25542	hypothetical prote
2	36	85.7	332	2 B64184	oligopeptide trans
3	36	85.7	745	2 C83011	hypothetical prote
4	35	83.3	91	2 T06008	hypothetical prote
5	35	83.3	104	2 H85303	hypothetical prote
6	33	78.6	194	2 JCT685	Mev17-like protein
7	33	78.6	205	2 AC2766	conserved hypothet
8	33	78.6	205	2 G37546	hypothetical prote
9	33	78.6	229	1 Q5AEC	replication initia
10	33	78.6	257	2 T29183	hypothetical prote
11	33	78.6	265	2 D83577	conserved hypothet
12	33	78.6	281	2 A30471	hypothetical prote
13	33	78.6	372	2 H70813	probable cysteine
14	33	78.6	585	2 S06958	spingomyelin phos
15	33	78.6	629	1 A39825	spingomyelin phos
16	33	78.6	637	2 F82218	conserved hypothet
17	33	78.6	677	2 S32816	potassium channel
18	33	78.6	1726	2 A39401	merozoite surface
19	32	76.2	231	2 T21119	hypothetical prote
20	32	76.2	319	2 AH2062	aldo/keto reductas
21	32	76.2	322	2 G70463	dolichol-phosphate
22	32	76.2	331	2 S05800	hypothetical prote
23	32	76.2	385	2 AH2798	UDP-glucose:cerami
24	32	76.2	387	2 H97577	ceramide glucosylt
25	32	76.2	398	2 I59399	oncogene PTF-1 - h
26	32	76.2	410	2 F88978	protein P37B4.4 (i
27	32	76.2	444	2 T15907	hypothetical prote
28	32	76.2	453	2 G83043	probable transport
29	32	76.2	461	1 EFRM51	translation elonga

C:Superfamily: inner membrane protein malk; ATP-binding cassette homology  
 F:39-240/Domain: ATP-binding cassette homology <ABC>  
 F:56-64/Region: nucleotide-binding motif A (P-loop)  
 F:184-168/Region: nucleotide-binding motif B

Query Match 85.7%; Score 36; DB 2; Length 332;  
 Best Local Similarity 71.4%; Pred. No. 10;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 WRQTRKD 7  
 DB 95 WKETRKD 101

RESULT 3  
 C83011  
 Hypothetical protein PA5089 [imported] - Pseudomonas aeruginosa (strain PAO1)  
 C:Species: Pseudomonas aeruginosa  
 C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
 R:Accession: C83011  
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B  
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lin,  
 ; Lory, S.; Olson, M.V.  
 Nature 406, 959-964, 2000  
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic patho  
 A:Reference number: A82950; MUID:20437337; PMID:10984043  
 A:Accession: C83011  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-745 <STO>  
 A:Cross-references: GB:AE004921; GB:AE004091; NID:99951372; PIDN:AAG08474.1; GSPDB:GN001  
 A:Experimental source: strain PAO1  
 C:Genetics:  
 A:Gene: PA5089

Query Match 85.7%; Score 36; DB 2; Length 745;  
 Best Local Similarity 85.7%; Pred. No. 23;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 WRQTRKD 7  
 DB 715 WRQTRKD 721

RESULT 4  
 T06008  
 Hypothetical protein T25K17.50 - Arabidopsis thaliana (fragment)  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 22-Oct-1999  
 C:Accession: T06008  
 R:Bevan, M.; Rose, M.; Hempel, S.; Entian, K.D.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X  
 submitted to the Protein Sequence Database, March 1999  
 A:Reference number: Z15382  
 A:Accession: T06008  
 A:Molecule type: DNA  
 A:Residues: 1-91 <BEV>  
 A:Cross-references: EMBL:AL049171; GSPDB:GN00062; ATSP:T25K17.50  
 A:Experimental source: cultivar Columbia; BAC clone T25K17  
 C:Genetics:  
 A:Gene: ATSP:T25K17.50  
 A:Map Position: 4  
 A:Introns: 37/1; 57/3

Query Match 83.3%; Score 35; DB 2; Length 91;  
 Best Local Similarity 71.4%; Pred. No. 4.5;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 WRQTRKD 7  
 DB 8 WRTRRD 14

RESULT 5  
 H85303  
 Hypothetical protein AT4g26240 [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 16-Feb-2001  
 C:Accession: H85303  
 R:anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spr  
 Nature 402, 769-777, 1999  
 A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.  
 A:Reference number: A85001; MUID:20083488; PMID:10617198  
 A:Accession: H85303  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-104 <STO>  
 A:Cross-references: GB:NC\_001268; NID:97269475; PIDN:CAB79479.1; GSPDB:GN00140  
 C:Genetics:  
 A:Gene: AT4g26240  
 A:Map Position: 4

Query Match 83.3%; Score 35; DB 2; Length 104;  
 Best Local Similarity 71.4%; Pred. No. 5.2;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 WRQTRKD 7  
 DB 8 WRTRRD 14

RESULT 6  
 JC7685  
 Mpv17-like protein, M-LP - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 30-Sep-2001  
 C:Accession: JC7685  
 R:Tida, R.; Yasuda, T.; Teubota, E.; Matsuki, T.; Kishi, K.  
 Biochem. Biophys. Res. Commun. 283, 292-296, 2001  
 A:Title: Cloning, mapping, genomic organization, and expression of mouse M-LP, a new m  
 A:Reference number: JC7685; MUID:21226134; PMID:11327696  
 A:Contents: C57BL/6  
 A:Accession: JC7685  
 A:Molecule type: mRNA  
 A:Residues: 1-194 <IID>  
 A:Cross-references: GB:AF305634  
 C:Comment: This protein, a new member of the peroxisomal membrane protein Mpv17 domain  
 mbane, and in transporting peroxisomal matrix components.  
 C:Genetics:  
 A:Gene: m-lp  
 A:Map position: 16B1-B2  
 A:Introns: 127/3; 137/3  
 C:Keywords: kidney; transmembrane protein

Query Match 78.6%; Score 33; DB 2; Length 194;  
 Best Local Similarity 83.3%; Pred. No. 25;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 WRQTRK 6  
 DB 45 WRQTRR 50

RESULT 7  
 AC2766  
 conserved hypothetical protein Atul540 [imported] - Agrobacterium tumefaciens (strain  
 C:Species: Agrobacterium tumefaciens  
 C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002  
 C:Accession: AC2766  
 R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo,  
 erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McCle  
 ; Karp, P.; Romero, P.; Zhang, S.  
 Science 294, 2317-2323, 2001  
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm  
 ster, E.W.  
 A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:Reference number: AB2577; MUID:21608550; PMID:11743193  
 A:Accession: AC2766  
 A:Status: Preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-205 <KUR>  
 A:Cross-references: GB:AE008698; PIDN:AAI42545.1; PID:g17739967; GSPDB:GN00186  
 A:Experimental source: strain C58 (Dupont)  
 C:Genetics:  
 A:Gene: Atul540  
 A:Map position: circular chromosome

Query Match 78.6%; Score 33; DB 2; Length 205;  
 Best Local Similarity 83.3%; Pred. No. 26;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WRQTRK 6  
 ||: |||  
 Db 36 WRERK 41

RESULT 8  
 G97546  
 hypothetical protein AGR\_C\_2837 [imported] - Agrobacterium tumefaciens (strain C58, Cere  
 C:Species: Agrobacterium tumefaciens  
 C:Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 18-Nov-2002  
 C:Accession: G97546  
 R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,  
 A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;  
 Science 294, 2323-2328, 2001  
 A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum  
 A:Reference number: A97359; MUID:21608551; PMID:11743194  
 A:Accession: G97546  
 A:Status: Preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-205 <KUR>  
 A:Cross-references: GB:AE007869; PIDN:AAK87328.1; PID:g15156626; GSPDB:GN00169  
 C:Genetics:  
 A:Gene: AGR\_C\_2837  
 A:Map position: circular chromosome

Query Match 78.6%; Score 33; DB 2; Length 205;  
 Best Local Similarity 83.3%; Pred. No. 26;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WRQTRK 6  
 ||: |||  
 Db 36 WRERK 41

RESULT 9  
 Q9SAEC  
 replication initiation protein E-229 - Staphylococcus aureus plasmid pC194  
 N:Contains: DNA ligase (EC 6.5.1.-); DNA lyase (EC 4.2.99.-)  
 C:Species: Staphylococcus aureus  
 C:Date: 18-Aug-1982 #sequence\_revision 18-Aug-1982 #text\_change 20-Apr-2000  
 C:Accession: A04491  
 R:Horinouchi, S.; Weisblum, B.  
 J. Bacteriol. 150, 815-825, 1982  
 A:Title: Nucleotide sequence and functional map of pC194, a plasmid that specifies induc  
 A:Reference number: A91791; MUID:82167188; PMID:6950931  
 A:Accession: A04491  
 A:Molecule type: DNA  
 A:Residues: 1-229 <HOR>  
 C:Genetics:  
 A:Gene: plasmid  
 C:Superfamily: Staphylococcus aureus plasmid pC194 hypothetical 27.5K protein  
 C:Keywords: carbon-oxygen lyase, ligase  
 F:214/Active site: Tyr (covalent DNA-binding) #status predicted

Query Match 78.6%; Score 33; DB 1; Length 229;  
 Best Local Similarity 71.4%; Pred. No. 29;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 WRQTRK 7  
 ||: |||  
 Db 70 WRKARKD 76

RESULT 10  
 T29183  
 hypothetical protein M6.3 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 31-Mar-2000  
 C:Accession: T29183  
 R:Minx, P.; Le, T.T.  
 submitted to the EMBL Data Library, April 1996  
 A:Description: The sequence of C. elegans cosmid M6.  
 A:Reference number: Z20584  
 A:Accession: T29183  
 A:Status: Preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-257 <MIN>  
 A:Cross-references: EMBL:U56860; PIDN:AAA98706.1; GSPDB:GN00028; CESP:M6.3  
 A:Experimental source: strain Bristol N2; clone M6  
 C:Genetics:  
 A:Gene: CESP:M6.3  
 A:Map position: X  
 A:Introns: 53/1; 65/1; 107/1; 210/3  
 C:Superfamily: unassigned homeobox proteins; homeobox homology  
 C:Keywords: DNA binding; homeobox; nucleus; transcription regulation

Query Match 78.6%; Score 33; DB 2; Length 257;  
 Best Local Similarity 71.4%; Pred. No. 33;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 WRQTRK 7  
 ||: |||  
 Db 222 WRRVRKD 228

RESULT 11  
 D83577  
 conserved hypothetical protein PA0550 [imported] - Pseudomonas aeruginosa (strain PA01)  
 C:Species: Pseudomonas aeruginosa  
 C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
 C:Accession: D83577  
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; E  
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lin  
 .; Lory, S.; Olson, M.V.  
 Nature 406, 959-964, 2000  
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pat  
 A:Reference number: A82950; MUID:20437337; PMID:10984043  
 A:Accession: D83577  
 A:Status: Preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-265 <STO>  
 A:Cross-references: GB:AE004491; GB:AE004091; NID:G9946412; PIDN:AA03939.1; GSPDB:GN00  
 A:Experimental source: strain PA01  
 C:Genetics:  
 A:Gene: PA0550  
 C:Superfamily: conserved hypothetical protein H1013

Query Match 78.6%; Score 33; DB 2; Length 265;  
 Best Local Similarity 71.4%; Pred. No. 34;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 WRQTRK 7  
 ||: |||  
 Db 258 WRATRRD 264

RESULT 12  
 A30471  
 hypothetical protein rep - Staphylococcus aureus plasmid pSK89  
 C:Species: Staphylococcus aureus  
 C:Date: 06-Dec-1991 #sequence\_revision 06-Dec-1991 #text\_change 20-Apr-2001

C/Accession: A30471  
 R/Littlejohn, T.G.; DiBerardino, D.; Messerotti, L.J.; Spiers, S.J.; Skurray, R.A.;  
 Gene 101, 59-66, 1991  
 A/Title: Structure and evolution of a family of genes encoding antiseptic and disinfectant  
 A/Reference number: A30471; PMID:91285435; PMID:1840534  
 A/Accession: A30471  
 A/Status: preliminary; not compared with conceptual translation  
 A/Molecule type: DNA  
 A/Residues: 1-281 <LIT>  
 C/Genetics:  
 A/Genome: plasmid  
 C/Superfamily: Staphylococcus aureus plasmid pC194 hypothetical 27.5K protein

Query Match 78.6%; Score 33; DB 2; Length 281;  
 Best Local Similarity 71.4%; Pred. No. 36;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 WRQTRKD 7  
 |||:  
 40 WRQTRKD 46

Db

RESULT 13  
 H70813  
 probable cysteine synthase - Mycobacterium tuberculosis (strain H37RV)  
 C/Species: Mycobacterium tuberculosis  
 C/Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 20-Jun-2000  
 C/Accession: H70813  
 R/Comor, R.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.;  
 Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Hoiroyd, S.;  
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A/Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
 A/Reference number: A70500; PMID:98295987; PMID:9634230  
 A/Accession: H70813  
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-372 <COL>  
 A/Cross-references: GB:AL022004; GB:AL123456; NID:G3261550; PIDN:CRA17654.1; PID:G291690  
 A/Experimental source: strain H37RV  
 C/Genetics:  
 A/Gene: cysM3  
 C/Superfamily: threonine dehydratase

Query Match 78.6%; Score 33; DB 2; Length 372;  
 Best Local Similarity 71.4%; Pred. No. 47;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 WRQTRKD 7  
 |||:  
 146 WQTRKD 152

Db

RESULT 14  
 S06958  
 sphingomyelin phosphodiesterase, acidic, inactive splice form 2 - human  
 N/Alternate names: acid sphingomyelinase, splice form 2  
 C/Species: Homo sapiens (man)  
 C/Date: 22-Jan-1993 #sequence\_revision 24-Nov-1999 #text\_change 21-Jan-2000  
 C/Accession: S06958; A39825  
 R/Quintern, L.E.; Schuchman, E.H.; Levran, O.; Suchi, M.; Ferlinz, K.; Reinke, H.; Sandh  
 EMO J. 8, 2469-2473, 1989  
 A/Title: Isolation of cDNA clones encoding human acid sphingomyelinase: occurrence of al  
 A/Reference number: S06957; PMID:90060003; PMID:2555181  
 A/Accession: S06958  
 A/Molecule type: mRNA  
 A/Residues: 270-585 <QUI>  
 A/Cross-references: EMBL:X52679; NID:G28881; PIDN:CRA36902.1; PID:G825629  
 R/Schuchman, E.H.; Suchi, M.; Takahashi, T.; Sandhoff, K.; Desnick, R.J.  
 J. Biol. Chem. 266, 8531-8539, 1991  
 A/Title: Human acid sphingomyelinase. Isolation, nucleotide sequence, and expression of  
 A/Reference number: A39825; PMID:91217097; PMID:1840600

A/Accession: A39825  
 A/Molecule type: mRNA  
 A/Residues: 1-362; IGGFVALSPYGLRLISLNMPCSRNFWLL', INSTDPAQQLQWLVELQAEDRGD', 375-58;  
 A/Cross-references: GB:MS9916; NID:G179094; PIDN:AAAS8377.1; PID:G179095  
 C/Comment: This form lacked catalytic activity when expressed in COS-1 cells. Its funct  
 C/Comment: For the active splice form 1, see PIR:A39825.  
 C/Genetics:  
 A/Gene: GDB:SMPD1  
 A/Cross-references: GDB:128144; OMIM:257200  
 A/Map position: 1p15.4-1p15.4  
 A/Introns: 104/3; 362/2; 375/3; 401/2; 450/1  
 C/Superfamily: acid sphingomyelinase; phosphoesterase core homology; saposin repeat hon  
 C/Keywords: alternative splicing  
 F:81-175/Domain: saposin repeat homology <SAP>  
 F:200-320/Domain: phosphoesterase core homology <PEC>

Query Match 78.6%; Score 33; DB 2; Length 585;  
 Best Local Similarity 71.4%; Pred. No. 74;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 WRQTRKD 7  
 |||:  
 285 WQTRKD 291

Db

RESULT 15  
 A39825  
 sphingomyelin phosphodiesterase (EC 3.1.4.12), acidic, splice form 1 precursor (validat  
 N/Alternate names: acid sphingomyelinase  
 C/Species: Homo sapiens (man)  
 C/Date: 20-Mar-1992 #sequence\_revision 07-Jul-1995 #text\_change 22-Mar-2002  
 C/Accession: S27009; S36357; A39825; JX0276; S06957; I55567; A42801; B42801; C42801; D4  
 R/Newzella, D.; Stoffel, W.  
 Biol. Chem. Hoppe-Seyler 373, 1233-1238, 1992  
 A/Title: Molecular cloning of the acid sphingomyelinase of the mouse and the organizati  
 A/Reference number: S27009; PMID:9183402; PMID:1292508  
 A/Accession: S27009  
 A/Molecule type: DNA  
 A/Residues: 1-629 <NEW>  
 A/Cross-references: EMBL:X63600; NID:G556808  
 R/Hofmann, K.  
 submitted to the EMBL Data Library, December 1991  
 A/Reference number: S36357  
 A/Accession: S36357  
 A/Molecule type: DNA  
 A/Residues: 1-321, 'T', 323-629 <HOP>  
 A/Cross-references: EMBL:X63600; NID:G556808; PIDN:CRA45145.1; PID:G556809  
 R/Schuchman, E.H.; Suchi, M.; Takahashi, T.; Sandhoff, K.; Desnick, R.J.  
 J. Biol. Chem. 266, 8531-8539, 1991  
 A/Title: Human acid sphingomyelinase. Isolation, nucleotide sequence, and expression o  
 A/Reference number: A39825; PMID:91217097; PMID:1840600  
 A/Accession: A39825  
 A/Molecule type: mRNA  
 A/Residues: 1-629 <SCH>  
 A/Cross-references: GB:MS9916; NID:G179094; PIDN:AAAS8377.1; PID:G179095  
 A/Note: polymorphisms were demonstrated at positions 322 and 506  
 R/Ida, H.; Renner, O.M.; Eto, Y.; Chan, W.Y.  
 J. Biochem. 114, 15-20, 1993  
 A/Title: Cloning of a human acid sphingomyelinase cDNA with a new mutation that renderi  
 A/Reference number: U0276; PMID:94012573; PMID:8407868  
 A/Accession: JX0276  
 A/Molecule type: mRNA  
 A/Residues: 1-35, 38-156, 'R', 158-321, 'T', 323-505, 'G', 507-629 <IDA>  
 A/Cross-references: EMBL:X59960; NID:G402620; PIDN:CAA42584.1; PID:G402621  
 A/Note: the authors translated the codon CTC for residue 85 as Thr and ACC for residue  
 A/Note: this form has no catalytic activity; correction of position 157 to Cys restore  
 R/Quintern, L.E.; Schuchman, E.H.; Levran, O.; Suchi, M.; Ferlinz, K.; Reinke, H.; San  
 EMO J. 8, 2469-2473, 1989  
 A/Title: Isolation of cDNA clones encoding human acid sphingomyelinase: occurrence of  
 A/Reference number: S06957; PMID:90060003; PMID:2555181  
 A/Accession: S06957  
 A/Molecule type: mRNA  
 A/Residues: 128-629 <QUI>

A;Cross-references: EMBL:X52678; NID:g28879; PIDN:CAA36901.1; PID:g28880  
 A;Note: parts of this sequence were confirmed by peptide sequencing  
 R;Levrin, O.; Desnick, R.J.; Schuchman, E.H.  
 J. Clin. Invest. 88, 806-810, 1991  
 A;Title: Niemann-Pick type B disease. Identification of a single codon deletion in the a  
 A;Reference number: I55567; MUID:91358737; PMID:1885770  
 A;Accession: I55567  
 A;Status: translation not shown; translated from GB/EMBL/DDBJ  
 A;Molecule type: DNA  
 A;Residues: 119-120, 'H', 122 <LEV>  
 A;Cross-references: GB:S55766; NID:g234719; PIDN:AAB19680.1; PID:g234720  
 R;Takahashi, T.; Suchi, M.; Desnick, R.J.; Takada, G.; Schuchman, E.H.  
 J. Biol. Chem. 267, 12552-12558, 1992  
 A;Title: Identification and expression of five mutations in the human acid sphingomyelin  
 opathic and non-neuropathic forms.  
 A;Reference number: A42801; MUID:92316934; PMID:1618760  
 A;Contents: annotation; characterization of mutations  
 A;Note: substitution of Ile for 382-Met or Ser for 383-Asn result in complete inactivati  
 tion and lead to milder (type B) disease  
 C;Comment: Two isoforms, neutral and acidic, have been identified. The acidic isoform is  
 C;Comment: For the inactive splice form 2, see PIR:S05958.  
 C;Genetics:  
 A;Gene: GDB:SMPD1  
 A;Cross-references: GDB:128144; OMIM:257200  
 A;Map position: 11p15.4-11p15.4  
 A;Introns: 104/3; 362/2; 419/3; 445/2; 494/1  
 A;Note: a defect in this gene may result in Niemann-Pick disease  
 C;Complex: monomer  
 C;Function:  
 A;Description: catalyzes the hydrolysis of sphingomylin to form phosphorylcholine and ce  
 C;Superfamily: acid sphingomyelinase; phosphoesterase core homology; saposin repeat hom  
 C;Keywords: acetylated amino end; alternative splicing; glycoprotein; lysosome; metallo  
 F:1-48/Domain: signal sequence #status predicted <SIG>  
 F:49-629/Product: sphingomyelin phosphodiesterase #status predicted <MAT>  
 F:81-175/Domain: saposin repeat homology <SAP>  
 F:200-320/Domain: phosphoesterase core homology <PEC>  
 F:49/Modified site: acetylated amino end (Asp) (in mature form) #status predicted  
 F:86,175,335,395,503,520/Binding site: carbohydrate (Asn) (covalent) #status predicted  
 F:89-165,92-157,120-131/Disulfide bonds: #status predicted  
 F:251,319,439/Active site: Asp, His, Tyr #status predicted

Query Match 78.6%; Score 33; DB 1; Length 629;  
 Best Local Similarity 71.4%; Pred. No. 80;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 WRQTRKD 7  
 |||||  
 Db 285 WHQTRQD 291

Search completed: February 11, 2004, 17:11:31  
 Job time : 27.3333 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 17:09:56 / Search time 24.6667 Seconds  
(without alignments)  
59.419 Million cell updates/sec

Title: US-09-901-187C-1

Perfect score: 42  
Sequence: 1 WRQTRKD 7

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 801455 seqs, 20932283 residues

Total number of hits satisfying chosen parameters: 801455

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: Published Applications AA:

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09D\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10D\_PUBCOMB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	100.0	7	10	US-09-901-187B-1
2	42	100.0	11	9	US-09-873-676-23
3	36	85.7	63	15	US-10-083-357-1168
4	34	81.0	542	12	US-10-017-161-2024
5	34	81.0	542	12	US-10-292-798-1670
6	34	81.0	783	12	US-10-094-749-1772
7	33	78.6	103	12	US-10-029-386-32093
8	33	78.6	147	12	US-10-108-260A-4016
9	33	78.6	153	12	US-10-002-631C-92
10	33	78.6	472	12	US-10-205-219-149
11	33	78.6	567	12	US-10-094-749-2480
12	33	78.6	629	12	US-10-352-762-2
13	33	78.6	1835	12	US-09-935-541-5
14	33	78.6	1835	12	US-10-425-800-5
15	32	76.2	62	12	US-09-864-408A-1290

16	32	76.2	65	15	US-10-156-761-9796	Sequence 9796, Ap
17	32	76.2	121	12	US-10-364-049-3441	Sequence 3441, Ap
18	32	76.2	142	12	US-10-364-049-3522	Sequence 3522, Ap
19	32	76.2	196	12	US-10-364-049-4324	Sequence 4324, Ap
20	32	76.2	199	10	US-09-796-692-1216	Sequence 1216, Ap
21	32	76.2	199	10	US-09-796-692-1675	Sequence 1675, Ap
22	32	76.2	199	10	US-09-796-692-2234	Sequence 2234, Ap
23	32	76.2	199	12	US-10-057-475B-1216	Sequence 1216, Ap
24	32	76.2	199	12	US-10-057-475B-1675	Sequence 1675, Ap
25	32	76.2	199	12	US-10-057-475B-2234	Sequence 2234, Ap
26	32	76.2	199	12	US-10-154-884B-1216	Sequence 1216, Ap
27	32	76.2	199	12	US-10-154-884B-1675	Sequence 1675, Ap
28	32	76.2	199	12	US-10-154-884B-2234	Sequence 2234, Ap
29	32	76.2	199	15	US-10-040-862-1216	Sequence 1216, Ap
30	32	76.2	199	15	US-10-040-862-1675	Sequence 1675, Ap
31	32	76.2	199	15	US-10-040-862-2234	Sequence 2234, Ap
32	32	76.2	277	15	US-10-156-761-11584	Sequence 11584, A
33	32	76.2	299	12	US-10-369-493-10700	Sequence 10700, A
34	32	76.2	343	9	US-09-864-761-36084	Sequence 36084, A
35	32	76.2	462	12	US-10-374-979-97	Sequence 97, Appl
36	32	76.2	523	12	US-10-369-493-12617	Sequence 12617, A
37	32	76.2	532	12	US-10-369-493-8540	Sequence 8540, Ap
38	32	76.2	614	10	US-09-331-631A-21	Sequence 21, Appl
39	32	76.2	614	11	US-09-847-208-27	Sequence 27, Appl
40	32	76.2	614	12	US-10-147-095-21	Sequence 21, Appl
41	32	76.2	614	12	US-10-100-303A-8	Sequence 8, Appl
42	32	76.2	626	11	US-09-847-208-28	Sequence 28, Appl
43	32	76.2	626	12	US-10-100-303A-7	Sequence 7, Appl
44	32	76.2	626	12	US-10-245-871-10	Sequence 10, Appl
45	32	76.2	626	15	US-10-228-806-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1

US-09-901-187B-1  
; Sequence 1, Application US/09901187B  
; Patent No. US20020151464A1  
; GENERAL INFORMATION:  
; APPLICANT: Paraceta Pharmaceuticals, Inc.  
; APPLICANT: Wolozin, Benjamin  
; APPLICANT: Osetrova-Golts, Natalie  
; APPLICANT: Lebowitz, Michael S.  
; TITLE OF INVENTION: Methods for Preventing Neural Tissue Damage and for the Treatment of  
; TITLE OF INVENTION: Alpha-Synuclein Diseases  
; FILE REFERENCE: PAN01/002US  
; CURRENT APPLICATION NUMBER: US/09/901,187B  
; CURRENT FILING DATE: 2001-07-09  
; PRIOR APPLICATION NUMBER: US 60/217,319  
; PRIOR FILING DATE: 2000-07-07  
; PRIOR APPLICATION NUMBER: US 60/279,199  
; PRIOR FILING DATE: 2001-03-28  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 7  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-901-187B-1

Query Match 100.0%; Score 42; DB 10; Length 7;  
Best Local Similarity 100.0%; Pred. No. 7e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WRQTRKD 7  
| | | | |  
Db 1 WRQTRKD 7

RESULT 2

US-09-873-676-23  
; Sequence 23, Application US/09873676

Patent No. US20020077289A1  
; GENERAL INFORMATION:  
; APPLICANT: MacDonald, Nicholas J.  
; APPLICANT: Sim, Kim L.  
; TITLE OF INVENTION: Angiotensin and Endostatin Binding Proteins and Methods of Use  
; FILE REFERENCE: 05213-0378 (43170-259333)  
; CURRENT FILING DATE: 2001-06-04  
; PRIOR APPLICATION NUMBER: US 60/209,065  
; PRIOR FILING DATE: 2000-06-02  
; PRIOR APPLICATION NUMBER: US 60/289,387  
; PRIOR FILING DATE: 2001-05-08  
; NUMBER OF SEQ ID NOS: 123  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 23  
; LENGTH: 11  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: synthetic binding peptide  
US-09-873-676-23

Query Match 100.0%; Score 42; DB 9; Length 11;  
Best Local Similarity 100.0%; Pred. No. 0.44; 0; Indels 0; Gaps 0;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WRQTRKD 7  
Db 1 WRQTRKD 7

RESULT 3  
US-10-083-357-1168  
; Sequence 1168, Application US/10083357  
; Publication No. US20030054370A1  
; GENERAL INFORMATION:  
; APPLICANT: Qianlong Zeng et al.  
; TITLE OF INVENTION: Systemic Discovery of New Genes  
; FILE REFERENCE: 032796-090  
; CURRENT APPLICATION NUMBER: US/10/083,357  
; CURRENT FILING DATE: 2002-02-27  
; NUMBER OF SEQ ID NOS: 1346  
; SEQ ID NO 1168  
; LENGTH: 63  
; TYPE: PRT  
; ORGANISM: Saccharomyces cerevisiae  
US-10-083-357-1168

Query Match 85.7%; Score 36; DB 15; Length 63;  
Best Local Similarity 85.7%; Pred. No. 24;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 WRQTRKD 7  
Db 18 WRQTRKD 24

RESULT 4  
US-10-017-161-2024  
; Sequence 2024, Application US/10017161  
; Publication No. US20030143668A1  
; GENERAL INFORMATION:  
; APPLICANT: SUMA, MAKIKO  
; APPLICANT: ASAI, KIYOSHI  
; APPLICANT: AKIYAMA, YUTAKA  
; APPLICANT: ABURATANI, HIROYUKI  
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS  
; FILE REFERENCE: 084335/0152  
; CURRENT APPLICATION NUMBER: US/10/017,161  
; CURRENT FILING DATE: 2002-12-18  
; PRIOR APPLICATION NUMBER: JP 2001/246789  
; PRIOR FILING DATE: 2001-06-18  
; NUMBER OF SEQ ID NOS: 2430

SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2024  
; LENGTH: 542  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-017-161-2024  
Query Match 81.0%; Score 34; DB 12; Length 542;  
Best Local Similarity 71.4%; Pred. No. 3.6e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WRQTRKD 7  
Db 5 WRQTRKD 11

RESULT 5  
US-10-292-798-1670  
; Sequence 1670, Application US/10292798  
; Publication No. US20030235833A1  
; GENERAL INFORMATION:  
; APPLICANT: SUMA, MAKIKO  
; APPLICANT: ASAI, KIYOSHI  
; APPLICANT: AKIYAMA, YUTAKA  
; APPLICANT: ABURATANI, HIROYUKI  
; TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS  
; FILE REFERENCE: 084335/166  
; CURRENT APPLICATION NUMBER: US/10/292,798  
; CURRENT FILING DATE: 2002-11-13  
; PRIOR APPLICATION NUMBER: 10/017,161  
; PRIOR FILING DATE: 2001-12-18  
; PRIOR APPLICATION NUMBER: JP 2001-246789  
; PRIOR FILING DATE: 2001-06-18  
; NUMBER OF SEQ ID NOS: 2070  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1670  
; LENGTH: 542  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-292-798-1670

Query Match 81.0%; Score 34; DB 12; Length 542;  
Best Local Similarity 71.4%; Pred. No. 3.6e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WRQTRKD 7  
Db 5 WRQTRKD 11

RESULT 6  
US-10-094-749-1772  
; Sequence 1772, Application US/10094749  
; Publication No. US20030219741A1  
; GENERAL INFORMATION:  
; APPLICANT: ISOGAI, TAKAO  
; APPLICANT: SUGIYAMA, TOMOYASU  
; APPLICANT: OTSUKI, TETSUJI  
; APPLICANT: WAKAMATSU, AI  
; APPLICANT: SATO, HIROYUKI  
; APPLICANT: ISHII, SHIZUKO  
; APPLICANT: YAMAMOTO, JUN-ICHI  
; APPLICANT: ISONO, YUUKO  
; APPLICANT: HIO, YUKI  
; APPLICANT: OTSUKA, KAORU  
; APPLICANT: NAGAI, KEIICHI  
; APPLICANT: IRIE, RYOTARO  
; APPLICANT: TAMECHIKA, ICHIRO  
; APPLICANT: SEKI, NAOHICO  
; APPLICANT: YOSHIKAWA, TSUTOMU  
; APPLICANT: OTSUKA, MOTOKUJI  
; APPLICANT: NAGAHARI, KENJI  
; APPLICANT: MASUHO, YASUHIKO



```

; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1772
; LENGTH: 783
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-094-749-1772

Query Match      81.0%; Score 34; DB 12; Length 783;
Best Local Similarity 71.4%; Pred. No. 5e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy      1 WRQTRKD 7
Db      246 WRKTRED 252

RESULT 7
US-10-029-386-32093
; Sequence 32093, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 32093
; LENGTH: 103
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC020757.2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.51
; OTHER INFORMATION: SWISSPROT HIT: P39210, EVALUATE 3.00e-07
US-10-029-386-32093

Query Match      78.6%; Score 33; DB 12; Length 103;
Best Local Similarity 83.3%; Pred. No. 1.2e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      1 WRQTRK 6
Db      45 WRQTRR 50

RESULT 8
US-10-108-260A-4016
; Sequence 4016, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20040005560A1e1 full length cdna
; FILE REFERENCE: HI-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27

; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1772
; LENGTH: 783
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-094-749-1772

Query Match      78.6%; Score 33; DB 12; Length 147;
Best Local Similarity 83.3%; Pred. No. 1.7e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      1 WRQTRK 6
Db      45 WRQTRR 50

RESULT 9
US-10-002-631C-92
; Sequence 92, Application US/10002631C
; Publication No. US20030157486A1
; GENERAL INFORMATION:
; APPLICANT: Graff, Jonathon M.
; APPLICANT: Muenster, Matthew
; TITLE OF INVENTION: METHODS TO IDENTIFY SIGNAL SEQUENCES
; FILE REFERENCE: A34943 090495.0243
; CURRENT APPLICATION NUMBER: US/10/002,631C
; CURRENT FILING DATE: 2001-10-31
; PRIOR APPLICATION NUMBER: 60/300,309
; PRIOR FILING DATE: 2001-06-21
; NUMBER OF SEQ ID NOS: 324
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 92
; LENGTH: 153
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-002-631C-92

Query Match      78.6%; Score 33; DB 12; Length 153;
Best Local Similarity 83.3%; Pred. No. 1.8e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      1 WRQTRK 6
Db      11 WRQTRR 16

RESULT 10
US-10-205-219-149
; Sequence 149, Application US/10205219
; Publication No. US20030138803A1
; GENERAL INFORMATION:
; APPLICANT: Warner-Lambert Company
; APPLICANT: Lee, Kevin
; APPLICANT: Dixon, Alistair
; APPLICANT: Brooksbank, Robert
; APPLICANT: Pinnock, Robert
; TITLE OF INVENTION: Identification and Use of Molecules Implicated in Pain
; FILE REFERENCE: WL-A-018200
; CURRENT APPLICATION NUMBER: US/10/205,219
; CURRENT FILING DATE: 2002-07-24
; PRIOR APPLICATION NUMBER: GB 0118354.0
; PRIOR FILING DATE: 2001-07-27
; NUMBER OF SEQ ID NOS: 197
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 149
; LENGTH: 472
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: NY-REN-58
US-10-205-219-149
```

```

Query Match      78.6%; Score 33; DB 12; Length 472;
Best Local Similarity 71.4%; Pred. No. 4.8e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 WRQTRKD 7
      |||: ||
Db      289 WRQSEKD 295

RESULT 11
US-10-094-749-2480
; Sequence 2480, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAGHIKO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOKYU
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2480
; LENGTH: 567
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-2480

Query Match      78.6%; Score 33; DB 12; Length 567;
Best Local Similarity 71.4%; Pred. No. 5.7e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 WRQTRKD 7
      |||: ||
Db      418 WRQSEKD 424

RESULT 12
US-10-352-762-2
; Sequence 2, Application US/10352762
; Publication No. US20030153739A1
; GENERAL INFORMATION:
; APPLICANT: Schuchman, Edward H.
; APPLICANT: Desnick, Robert J.
; TITLE OF INVENTION: The Acid Sphingomyelinase Gene and
; Diagnosis of Niemann-Pick Disease
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.

```

```

; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/352,762
; FILING DATE: 28-Jan-2003
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/106,375
; FILING DATE: 28-Jun-1998
; APPLICATION NUMBER: US 07/695,472
; FILING DATE: 03-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Mirock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 6923-014
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 7908864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 629 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: Protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-10-352-762-2

Query Match      78.6%; Score 33; DB 12; Length 629;
Best Local Similarity 71.4%; Pred. No. 6.2e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 WRQTRKD 7
      |||: ||
Db      285 WHQTRQD 291

RESULT 13
US-09-935-541-5
; Sequence 5, Application US/09935541
; Patent No. US2002015091A1
; GENERAL INFORMATION:
; APPLICANT: Dietrich, Paul S.
; APPLICANT: McGivern, Joseph G.
; TITLE OF INVENTION: T-TYPE CALCIUM CHANNEL VARIANTS; COMPOSITIONS THEREOF;
; TITLE OF INVENTION: AND USES
; FILE REFERENCE: R0043B-REG sequence listing
; CURRENT APPLICATION NUMBER: US/09/935,541
; CURRENT FILING DATE: 2001-08-23
; PRIOR APPLICATION NUMBER: 09/404,650
; PRIOR FILING DATE: 1999-09-23
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 1835
; TYPE: PRT
; ORGANISM: Rattus sp.
US-09-935-541-5

Query Match      78.6%; Score 33; DB 10; Length 1835;
Best Local Similarity 83.3%; Pred. No. 1.6e+03;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 WRQTRK 6
      |||: ||
Db      584 WRQTRK 589

RESULT 14

```

```

US-10-425-800-5
; Sequence 5, Application US/10425800
; Publication No. US20030180886A1
; GENERAL INFORMATION:
; APPLICANT: Dietrich, Paul S.
; APPLICANT: McGiven, Joseph G.
; TITLE OF INVENTION: T-TYPE CALCIUM CHANNEL VARIANTS; COMPOSITIONS THEREOF;
; TITLE OF INVENTION: AND USES
; FILE REFERENCE: R0043B-REG sequence listing
; CURRENT APPLICATION NUMBER: US/10/425,800
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US/03/404,650
; PRIOR FILING DATE: 1999-09-23
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 1835
; TYPE: PRT
; ORGANISM: Rattus sp.
US-10-425-800-5

Query Match      78.6%; Score 33; DB 12; Length 1835;
Best Local Similarity 83.3%; Pred. No. 1.6e+03;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      1 WRQTRK 6
Db      584 WRETRK 589

RESULT 15
US-09-864-408A-1290
; Sequence 1290, Application US/09864408A
; Publication No. US20040009474A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Shimkets, Richard A.
; FILE REFERENCE: 21402-012
; CURRENT APPLICATION NUMBER: US/09/864,408A
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 60/206,690
; PRIOR FILING DATE: 2000-05-24
; NUMBER OF SEQ ID NOS: 9068
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1290
; LENGTH: 62
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-864-408A-1290

Query Match      76.2%; Score 32; DB 12; Length 62;
Best Local Similarity 83.3%; Pred. No. 1.2e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      1 WRQTRK 6
Db      56 WQSRK 61

Search completed: February 11, 2004, 17:54:03
Job time : 25.6667 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 16:56:09 ; Search time 10.4167 Seconds  
(without alignments)  
28.433 Million cell updates/sec

Title: US-09-901-187C-1

Perfect score: 42

Sequence: 1 WRQTRKD 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA.\*

- 1: /cgn2\_6/ptodata/1/iaa/5A.COMB.pep.\*
- 2: /cgn2\_6/ptodata/1/iaa/5B.COMB.pep.\*
- 3: /cgn2\_6/ptodata/1/iaa/6A.COMB.pep.\*
- 4: /cgn2\_6/ptodata/1/iaa/6B.COMB.pep.\*
- 5: /cgn2\_6/ptodata/1/iaa/PCFUS.COMB.pep.\*
- 6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	100.0	11	3	US-09-206-059-23
2	36	85.7	341	4	US-09-252-991A-32629
3	36	85.7	748	4	US-09-252-991A-32942
4	35	83.3	435	4	US-09-252-991A-19756
5	33	78.6	293	4	US-09-252-991A-19886
6	33	78.6	428	4	US-09-252-991A-19297
7	33	78.6	472	4	US-09-166-350-17
8	33	78.6	629	1	US-08-250-740-33
9	33	78.6	629	1	US-07-695-472B-2
10	33	78.6	629	4	US-09-106-375-2
11	33	78.6	1835	4	US-09-404-650-5
12	32	76.2	114	4	US-09-183-953-14
13	32	76.2	224	4	US-09-347-650-2
14	32	76.2	397	2	US-08-371-377-19
15	32	76.2	398	2	US-08-371-377-17
16	32	76.2	462	1	US-08-299-351-1
17	32	76.2	462	2	US-08-371-377-18
18	32	76.2	462	4	US-09-309-572-10
19	32	76.2	462	6	525348-1
20	32	76.2	474	4	US-09-252-991A-26967
21	32	76.2	626	4	US-09-106-872A-4
22	32	76.2	787	4	US-09-252-991A-19991
23	32	76.2	1337	3	US-08-854-585-2
24	32	76.2	1337	4	US-09-447-533-2
25	32	76.2	1337	5	PCT-US95-05512-2
26	31	73.8	123	4	US-09-615-192A-309
27	31	73.8	179	4	US-09-615-192A-289

Sequence 29093, A  
Sequence 30066, A  
Sequence 21276, A  
Sequence 21050, A  
Sequence 20, Appl  
Sequence 7, Appl  
Sequence 10, Appl  
Sequence 343, App  
Sequence 4, Appl  
Sequence 4, Appl  
Sequence 4, Appl  
Sequence 11, Appl  
Sequence 311, App  
Sequence 11, Appl  
Sequence 4772, Ap  
Sequence 30, Appl  
Sequence 6, Appl  
Sequence 9, Appl  
Sequence 4, Appl

US-09-252-991A-29093  
US-09-252-991A-30066  
US-09-252-991A-21276  
US-09-252-991A-21050  
US-09-126-420A-20  
US-09-499-302A-7  
US-09-499-302A-10  
US-09-198-452A-343  
US-08-117-362-4  
US-08-486-924-4  
US-08-486-929A-4  
US-08-615-192A-311  
US-09-100-804-11  
US-09-328-352-4772  
US-08-313-075A-30  
US-09-292-768-6  
US-08-881-784-9  
US-09-292-768-4

ALIGNMENTS

RESULT 1

US-09-206-059-23  
; Sequence 23, Application US/09206059  
; Patent No. 6201104  
; GENERAL INFORMATION:  
; APPLICANT: MacDonald, Nicholas  
; APPLICANT: Sim, Kim Lee  
; TITLE OF INVENTION: Angiogenesis-Inhibiting Protein Binding Peptides and  
; TITLE OF INVENTION: Proteins and Methods of Use  
; FILE REFERENCE: 05213-0370  
; CURRENT APPLICATION NUMBER: US/09/206,059  
; CURRENT FILING DATE: 1998-12-04  
; NUMBER OF SEQ ID NOS: 80  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 23  
; LENGTH: 11  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: binding peptides  
US-09-206-059-23

Query Match 100.0%; Score 42; DB 3; Length 11.  
Best Local Similarity 100.0%; Pred. No. 0.031; 0; Gaps 0;  
Matches 7; Conservative 0; Mismatches 0; Indels 0

QY 1 WRQTRKD 7

Db 1 WRQTRKD 7

RESULT 2

US-09-252-991A-32629  
; Sequence 32629, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 32629  
; LENGTH: 341

```
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32629

Query Match      85.7%; Score 36; DB 4; Length 341;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 WRQTRK 6
      |||||
Db      17 WRQTRK 22

RESULT 3
US-09-252-991A-32942
; Sequence 32942, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 32942
; LENGTH: 748
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32942

Query Match      85.7%; Score 36; DB 4; Length 748;
Best Local Similarity 85.7%; Pred. No. 31;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 WRQTRKD 7
      |||||
Db      718 WRQTRKD 724

RESULT 4
US-09-252-991A-19756
; Sequence 19756, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 19756
; LENGTH: 435
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19756

Query Match      83.3%; Score 35; DB 4; Length 435;
Best Local Similarity 71.4%; Pred. No. 28;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 WRQTRKD 7
      |||||
Db      392 WRRTRRD 398
```

```
RESULT 5
US-09-252-991A-19896
; Sequence 19896, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 19896
; LENGTH: 293
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19896

Query Match      78.6%; Score 33; DB 4; Length 293;
Best Local Similarity 71.4%; Pred. No. 46;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 WRQTRKD 7
      |||||
Db      286 WRRTRRD 292

RESULT 6
US-09-252-991A-19297
; Sequence 19297, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 19297
; LENGTH: 428
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19297

Query Match      78.6%; Score 33; DB 4; Length 428;
Best Local Similarity 83.3%; Pred. No. 68;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 WRQTRK 6
      |||||
Db      266 WRETRK 271

RESULT 7
US-09-166-350-17
; Sequence 17, Application US/09166350A
; Patent No. 6440663
; GENERAL INFORMATION:
; APPLICANT: Scanlan, Matthew
; APPLICANT: Chen, Yao
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Old, Lloyd
```

; APPLICANT: Jager, Elke  
; APPLICANT: Knuth, Alex  
; TITLE OF INVENTION: Renal Cancer Associated Antigens and  
; TITLE OF INVENTION: Uses Therefor  
; FILE REFERENCE: L0461/7051  
; CURRENT APPLICATION NUMBER: US/09/166,350A  
; CURRENT FILING DATE: 1998-10-05  
; EARLIER APPLICATION NUMBER: US 09/166,350  
; EARLIER FILING DATE: 1998-10-05  
; NUMBER OF SEQ ID NOS: 35  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 17  
; LENGTH: 472  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-166-350-17

Query Match 78.6%; Score 33; DB 4; Length 472;  
Best Local Similarity 71.4%; Pred. No. 75;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 WRQTRKD 7  
Db 289 WRQSEKD 295

RESULT 8  
US-08-250-740-33  
; Sequence 33, Application US/08250740  
; Patent No. 5686240  
; GENERAL INFORMATION:  
; APPLICANT: Schuchman, Edward H.  
; APPLICANT: Desnick, Robert J.  
; TITLE OF INVENTION: Acid Sphingomyelinase Gene and Diagnosis  
; TITLE OF INVENTION: of Niemann-Pick Disease  
; NUMBER OF SEQUENCES: 36  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/250,740  
; FILING DATE: 27-MAY-1994

; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A.  
; REGISTRATION NUMBER: 30742  
; REFERENCE/DOCKET NUMBER: 6923-038  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-8864  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 33:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 629 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-08-250-740-33

Query Match 78.6%; Score 33; DB 1; Length 629;  
Best Local Similarity 71.4%; Pred. No. 1e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 WRQTRKD 7

Db 285 WHQTRQD 291

RESULT 9  
US-07-695-472B-2  
; Sequence 2, Application US/07695472B  
; Patent No. 5773278  
; GENERAL INFORMATION:  
; APPLICANT: Schuchman, Edward H.  
; APPLICANT: Desnick, Robert J.  
; TITLE OF INVENTION: The Acid Sphingomyelinase Gene and  
; TITLE OF INVENTION: Diagnosis of Niemann-Pick Disease  
; NUMBER OF SEQUENCES: 36  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/695,472B  
; FILING DATE: 19910503  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mistrock, S. Leslie  
; REGISTRATION NUMBER: 18,872  
; REFERENCE/DOCKET NUMBER: 6923-014  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 7908864/9741  
; TELEX: 66141 PENNIE

; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 629 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-07-695-472B-2

Query Match 78.6%; Score 33; DB 1; Length 629;  
Best Local Similarity 71.4%; Pred. No. 1e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 WRQTRKD 7  
Db 285 WHQTRQD 291

RESULT 10  
US-09-106-375-2  
; Sequence 2, Application US/09106375  
; Patent No. 6541218  
; GENERAL INFORMATION:  
; APPLICANT: Schuchman, Edward H.  
; APPLICANT: Desnick, Robert J.  
; TITLE OF INVENTION: The Acid Sphingomyelinase Gene and  
; TITLE OF INVENTION: Diagnosis of Niemann-Pick Disease  
; NUMBER OF SEQUENCES: 36  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036  
; COMPUTER READABLE FORM:

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/106,375
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/695,472
; FILING DATE: 03-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Mirock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 6923-014
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 7908864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 629 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-09-106-375-2

Query Match 78.6%; Score 33; DB 4; Length 629;
Best Local Similarity 71.4%; Pred. No. 1e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 WRQTRKD 7
DB 285 WHQTRQD 291

RESULT 11
US-09-404-650-5
; Sequence 5, Application US/09404650
; Patent No. 6309858
; GENERAL INFORMATION:
; APPLICANT: Dietrich, Paul S.
; APPLICANT: McGilver, Joseph G.
; TITLE OF INVENTION: T-TYPE CALCIUM CHANNEL VARIANTS; COMPOSITIONS THEREOF;
; FILE REFERENCE: R0043B-REG sequence listing
; CURRENT APPLICATION NUMBER: US/09/404,650
; CURRENT FILING DATE: 1999-09-23
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; TYPE: PRT
; LENGTH: 1835
; ORGANISM: Rattus sp.
US-09-404-650-5

Query Match 78.6%; Score 33; DB 4; Length 1835;
Best Local Similarity 83.3%; Pred. No. 2.9e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WRQTRK 6
DB 584 WRQTRK 589

RESULT 12
US-09-183-959-14
; Sequence 14, Application US/09183959
; Patent No. 6303322
; GENERAL INFORMATION:
; APPLICANT: Cahoon, Rebecca E.
; APPLICANT: Miao, Guo-Hua
; APPLICANT: Rafalski, J. Antoni
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/183,959
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/695,472
; FILING DATE: 03-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Mirock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 6923-014
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 7908864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 629 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-09-183-959-14

Query Match 76.2%; Score 32; DB 4; Length 114;
Best Local Similarity 71.4%; Pred. No. 28;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 WRQTRKD 7
DB 102 WRQTRKD 108

RESULT 13
US-09-347-650-2
; Sequence 2, Application US/09347650
; Patent No. 6576814
; GENERAL INFORMATION:
; APPLICANT: Briggs, Steven
; TITLE OF INVENTION: Manipulation of Mlo Genes to Enhance Disease Resistance
; FILE REFERENCE: 5718-42035718/158714
; CURRENT APPLICATION NUMBER: US/09/347,650
; CURRENT FILING DATE: 1999-07-06
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 224
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Mlo1
US-09-347-650-2

Query Match 76.2%; Score 32; DB 4; Length 224;
Best Local Similarity 71.4%; Pred. No. 55;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 WRQTRKD 7
DB 116 WRQTRKD 122

RESULT 14
US-08-371-377-19
; Sequence 19, Application US/08371377
; Patent No. 5851764
; GENERAL INFORMATION:
; APPLICANT: Fisher, Paul B.
; APPLICANT: Shen, Ruogian
; TITLE OF INVENTION: DEVELOPMENT OF DNA PROBES AND
; TITLE OF INVENTION: IMMUNOLOGICAL REAGENTS SPECIFIC FOR CELL SURFACE-EXPRESSED
; TITLE OF INVENTION: MOLECULES AND TRANSFORMATION-ASSOCIATED GENES
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10036
; COMPUTER READABLE FORM:

```

Thu Feb 12 09:07:26 2004

TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-371-377-17  
Query Match 76.2%; Score 32; DB 2; Length 398;  
Best Local Similarity 71.4%; Pred. No. 99;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 WRQTRKD 7  
Db 150 WKVTRKD 156  
Search completed: February 11, 2004, 17:13:32  
Job time : 11.4167 secs

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/371.377  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 0575/37590-B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0525  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 397 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-371-377-19  
Query Match 76.2%; Score 32; DB 2; Length 397;  
Best Local Similarity 71.4%; Pred. No. 98;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 WRQTRKD 7  
Db 150 WKVTRKD 156  
RESULT 15  
US-08-371-377-17  
Sequence 17, Application US/08371377  
Patent No. 5851764  
GENERAL INFORMATION:  
APPLICANT: Fisher, Paul B.  
TITLE OF INVENTION: DEVELOPMENT OF DNA PROBES AND  
TITLE OF INVENTION: IMMUNOLOGICAL REAGENTS SPECIFIC FOR CELL SURFACE-EXPRESSED  
TITLE OF INVENTION: MOLECULES AND TRANSFORMATION-ASSOCIATED GENES  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: United States of America  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/371,377  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 0575/37590-B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0525  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 398 amino acids  
TYPE: amino acid



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 16:51:33 ; Search time 32.25 Seconds  
(without alignments)  
34.452 Million cell updates/sec

Title: US-09-901-187C-1

Perfect score: 42

Sequence: 1 WRQTRKD 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_19Jun03.\*

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	42	100.0	7	AAE14546	Human alpha-synuc
2	42	100.0	11	AAE16472	Linear peptide tha
3	42	100.0	11	AAW48844	Endostatin protein
4	35	83.3	198	AAAG56017	Arabidopsis thalia
5	34	81.0	95	AAU55719	Propionibacterium
6	34	81.0	119	AAU65599	Propionibacterium
7	34	81.0	176	AAU50314	Propionibacterium
8	34	81.0	388	AAU11732	Human MDTT polypep
9	34	81.0	493	ABU11557	Human MDTT polypep

10	34	81.0	563	23	AAU98880	Human patched-like
11	33	78.6	381	18	AAW35283	Human acid sphingo
12	33	78.6	472	21	AAW22343	Human cancer assoc
13	33	78.6	628	14	AAW30644	DeltAR608 ASM. Ho
14	33	78.6	628	18	AAW35262	Human acid sphingo
15	33	78.6	629	14	AAW30642	ASM protein. Homo
16	33	78.6	629	14	AAW30643	R496L ASM. Homo s
17	33	78.6	629	14	AAW30645	L302P ASM. Homo s
18	33	78.6	629	18	AAW35260	Human acid sphingo
19	33	78.6	629	18	AAW35261	Human acid sphingo
20	33	78.6	629	18	AAW35263	Plant potassium ch
21	33	78.6	677	14	AAW43652	Angiotensin conver
22	33	78.6	682	22	AAU02902	Drosophila melanog
23	33	78.6	729	22	ABW68566	Rat brain alpha-II
24	33	78.6	1792	22	ABW66477	Rat T-type voltage
25	33	78.6	1835	20	AAW14597	Human ORF645 prote
26	32	76.2	62	23	ABP31672	Propionibacterium
27	32	76.2	66	22	AAU45060	Novel human diagno
28	32	76.2	110	22	ABG05483	Zea mays SCUBr pro
29	32	76.2	114	20	AAW14141	Human ovarian anti
30	32	76.2	121	23	ABP42309	Novel human secret
31	32	76.2	122	22	AAU32987	Human polypeptide
32	32	76.2	132	22	AAO03503	Zinc finger protei
33	32	76.2	139	23	ABW98922	Human ovarian anti
34	32	76.2	142	23	ABP43390	Human MDTT polypep
35	32	76.2	184	24	ABU11525	Human ovarian anti
36	32	76.2	196	23	ABP43192	Human haematologic
37	32	76.2	199	22	AAW80852	Human haematologic
38	32	76.2	199	22	AAW81311	Human haematologic
39	32	76.2	199	22	AAW81870	Drosophila melanog
40	32	76.2	216	22	ABW61225	Maize MLO1 protein
41	32	76.2	224	21	AAW44601	Human liver peptid
42	32	76.2	343	22	ABG50215	Peptide #2830 enco
43	32	76.2	343	22	ABW30179	Peptide #2851 enco
44	32	76.2	343	22	ABW35345	Protein #2785 enco
45	32	76.2	343	22	ABW20786	

ALIGNMENTS

RESULT 1

AAE14546

ID AAE14546 standard; peptide; 7 AA.

AC AAE14546;

DT 17-MAY-2002 (first entry)

DE Human alpha-synuclein aggregation inhibitor #1.

Human alpha-synuclein aggregation inhibitor #1.  
Alpha-synuclein; inhibitor; neurodegenerative disease; Lewy body;  
Parkinson's disease; Alzheimer's disease; diffuse Lewy body disease;  
multiple system atrophy; Hallervorden-Spatz disease; human.

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	42	100.0	7	AAE14546	Human alpha-synuc
2	42	100.0	11	AAE16472	Linear peptide tha
3	42	100.0	11	AAW48844	Endostatin protein
4	35	83.3	198	AAAG56017	Arabidopsis thalia
5	34	81.0	95	AAU55719	Propionibacterium
6	34	81.0	119	AAU65599	Propionibacterium
7	34	81.0	176	AAU50314	Propionibacterium
8	34	81.0	388	AAU11732	Human MDTT polypep
9	34	81.0	493	ABU11557	Human MDTT polypep

PT synuclein useful for treating a neurodegenerative disease involves  
 PT determining aggregation of alpha synuclein in the presence of exogenous  
 XX iron or copper  
 XX Claim 40; Page 37; 52pp; English.

XX The invention relates to screening of inhibitors of alpha-synuclein  
 CC aggregation in the presence of exogenous iron or copper. The inhibitors  
 CC are magnesium and alpha-synuclein binding peptides, which are  
 CC useful for treating neurodegenerative disease that involves  
 CC the formation of Lewy bodies e.g. Parkinson's disease (PD), Alzheimer's  
 CC disease (AD), diffuse Lewy body disease, mixed AD-PD, multiple system  
 CC atrophy and Hallervorden-Spatz disease. The present sequence is a  
 CC peptide that binds to C-terminal portion of human alpha-synuclein and  
 CC inhibits its aggregation.

XX Sequence 7 AA;

Query Match 100.0%; Score 42; DB 23; Length 7;  
 Best Local Similarity 100.0%; Pred. NO. 9.3e+05; Indels 0; Gaps 0;  
 Matches 7; Conservative 0; Mismatches 0;

QY 1 WRQTRKD 7  
 |||||  
 Db 1 WRQTRKD 7

# RESULT 2

AA16472  
 ID AAB16472 standard; Peptide; 11 AA.

AC AAB16472;

DT 27-OCT-2000 (first entry)

DE Linear peptide that binds to endostatin SEQ ID # 23.

XX Angiogenesis-inhibiting protein receptor; angiogenesis; angiostatin;  
 KW endostatin; plasminogen; laminin; treatment; wound healing; solid tumour;  
 KW psoriasis; scleroderma; myocardial angiogenesis; Crohn's disease;  
 KW cerebral collaterals; arteriovenous malformation; rubecosis; cancer;  
 KW diabetic retinopathy; arthritis; wound healing; peptic ulcer;  
 KW Helicobacter related disease; fracture; cat scratch fever.

XX Synthetic.

XX WO200032631-A2.

XX 08-JUN-2000.

XX 06-DEC-1999; 98WO-US28897.

XX 04-DEC-1998; 98US-0206059.

XX (ENTR-) ENTREMED INC.

XX Macdonald NJ, Sim KL;

XX WPI; 2000-412290/35.

XX New angiogenesis-inhibiting protein receptors, useful in methods for  
 PT treating diseases and processes that are mediated by angiogenesis, such  
 PT as solid tumours, psoriasis, scleroderma and myocardial angiogenesis -  
 XX Claim 7; Page 36; 100pp; English.

XX This invention relates to angiogenesis-inhibiting protein receptors, and  
 CC the DNA sequences encoding them. Angiogenesis is the generation of new  
 CC blood vessels into a tissue, and normally occurs in wound healing,  
 CC foetal and embryonal development and the formation of the corpus luteum,  
 CC endometrium and placenta. Angiostatin is a protein (see AAB16450 and  
 CC AAB16452) involved in angiogenesis, and has an amino acid sequence  
 CC similar to that of a plasminogen fragment (see murine plasminogen

CC AAB16490). Angiostatin has the ability to inhibit angiogenesis.  
 CC Endostatin is also an angiogenesis inhibiting protein (see AAB16451 and  
 CC AAB16452). Sequences AAB16451 and AAB16452 represent coding and protein  
 CC sequences of human laminin. Laminin is an angiostatin binding protein,  
 CC and some of the peptides of the invention share homology with regions of  
 CC laminin. Peptides AAB16452-B16521 (excluding AAB16490) are the  
 CC angiogenesis-inhibiting protein receptor fragments of the invention. The  
 CC peptides bind either angiostatin or endostatin and can be used in methods  
 CC for treating diseases and processes that are mediated by angiogenesis,  
 CC such as solid tumours, psoriasis, scleroderma, myocardial angiogenesis,  
 CC Crohn's disease, cerebral collaterals, arteriovenous malformations,  
 CC rubecosis, diabetic retinopathy, arthritis, wound healing, peptic ulcers,  
 CC Helicobacter related diseases, fractures, placental and cat scratch  
 CC fever. They are useful for the detection and prognosis of cancer. DNA  
 CC sequences A628204-A628241 encode the peptides of the invention.

XX Sequence 11 AA;

Query Match 100.0%; Score 42; DB 21; Length 11;  
 Best Local Similarity 100.0%; Pred. No. 0.17; Indels 0; Gaps 0;  
 Matches 7; Conservative 0; Mismatches 0;

QY 1 WRQTRKD 7  
 |||||  
 Db 1 WRQTRKD 7

# RESULT 3

AA48844  
 ID AAM48844 standard; Peptide; 11 AA.

AC AAM48844;

DT 04-APR-2002 (first entry)

DE Endostatin protein binding peptide SEQ ID NO: 23.

XX Human; angiostatin; endostatin; angiogenesis; cancer; metastasis;  
 KW psoriasis; scleroderma; Crohn's disease; corneal disease;  
 KW retinopathy; arthritis; wound healing; Helicobacter pylori; peptic ulcer;  
 KW gene therapy; angiostatin antagonist; endostatin antagonist;  
 KW antiangiogenic; cytostatic; antiarthritic; antiinflammatory;  
 KW cerebroprotective; antidiabetic; virucide; antipyretic; vulnary;  
 KW gynaecological; cat scratch fever.

XX Synthetic.

XX WO200193897-A2.

XX 13-DEC-2001.

XX 04-JUN-2001; 2001WO-US17947.

XX 02-JUN-2000; 2000US-209065P.

XX 08-MAY-2001; 2001US-289387P.

XX (ENTR-) ENTREMED INC.

XX Sim KL, Macdonald NJ;

XX WPI; 2002-130569/17.

XX Regulating angiogenesis and treatment of angiogenesis-mediated  
 PT diseases, e.g. haemangioma, tumors or cancer, by administering a  
 PT tropomyosin binding compound or actin disrupting compound -  
 XX Example 1; Page 44; 95pp; English.

XX The present invention relates to methods of regulating angiogenesis in an  
 CC individual by administering an angiogenesis regulating composition  
 CC comprising a tropomyosin binding compound or an actin disrupting  
 CC compound. The compositions are useful for treating diseases and processes  
 CC mediated by angiogenesis including haemangioma, solid tumours, blood

CC bourne tumours, leukaemia, metastasis, Crohn's disease, coronary or  
CC cerebral collaterals, arthritis, diabetic neovascularisation, macular  
CC degeneration, wound healing, Helicobacter related diseases, ovulation,  
CC mensturation, and cat scratch fever. The present sequence is a peptide  
CC described in the exemplification of the invention.

XX Sequence 11 AA

Query Match 100.0%; Score 42; DB 23; Length 11;  
Best Local Similarity 100.0%; Pred. No. 0.17; 0; Gaps 0;  
Matches 7; Conservative 0; Mismatches 0; Indels 0;

QY 1 WRQTRKD 7  
Db 1 WRQTRKD 7

## RESULT 4

AAG56017  
ID AAG56017 standard; Protein; 198 AA.

AC AAG56017;

XX 18-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 71932.

DE Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; Genetic mapping; Gene expression control; promoter;  
KW termination sequence.

XX Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-030439.

XX 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

XX 09-MAR-1999; 99US-0123548.

XX 23-MAR-1999; 99US-0125788.

XX 25-MAR-1999; 99US-0126264.

XX 29-MAR-1999; 99US-0126785.

XX 01-APR-1999; 99US-0127462.

XX 06-APR-1999; 99US-0128234.

XX 08-APR-1999; 99US-0128714.

XX 16-APR-1999; 99US-0129845.

XX 19-APR-1999; 99US-0130077.

XX 21-APR-1999; 99US-0130449.

XX 23-APR-1999; 99US-0130510.

XX 28-APR-1999; 99US-0130891.

XX 30-APR-1999; 99US-0131449.

XX 30-APR-1999; 99US-0132048.

XX 04-MAY-1999; 99US-0132407.

XX 05-MAY-1999; 99US-0132484.

XX 06-MAY-1999; 99US-0132485.

XX 06-MAY-1999; 99US-0132486.

XX 07-MAY-1999; 99US-0132487.

XX 11-MAY-1999; 99US-0132863.

XX 14-MAY-1999; 99US-0134256.

XX 14-MAY-1999; 99US-0134218.

XX 14-MAY-1999; 99US-0134219.

XX 14-MAY-1999; 99US-0134221.

XX 18-MAY-1999; 99US-0134370.

XX 19-MAY-1999; 99US-0134768.

XX 20-MAY-1999; 99US-0134941.

XX 21-MAY-1999; 99US-0135124.

XX 21-MAY-1999; 99US-0135353.

XX 24-MAY-1999; 99US-0135629.

XX 25-MAY-1999; 99US-0136021.

XX 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139452.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 21-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142330.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.



Query Match 81.0%; Score 34; DB 22; Length 95;  
Best Local Similarity 71.4%; Pred. No. 43;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 WRQTRKD 7  
||| |  
DB 8 WRQVRD 14

RESULT 6  
AAU6599  
ID AAU6599 standard; Protein; 119 AA.  
XX AC  
XX AAU6599;  
XX DT 27-FEB-2002 (first entry)  
XX DE Propionibacterium acnes immunogenic protein #27495.  
XX KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
XX KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
XX KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
XX KW dermatological; osteopathic; neuroprotectant.  
XX OS Propionibacterium acnes.  
XX PN WO200181581-A2.  
XX PD 01-NOV-2001.  
XX PF 20-APR-2001; 2001WO-US12865.  
XX PR 21-APR-2000; 2000US-199047P.  
XX PR 02-JUN-2000; 2000US-208841P.  
XX PR 07-JUL-2000; 2000US-216747P.  
XX PA (CORI-) CORIXA CORP.  
XX PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
XX PI L'maisonneuve J, Zhang Y, Jen S, Carter D;  
XX WPI; 2001-616774/71.  
XX DR N-PSDB; AAS59744.  
XX PT Propionibacterium acnes polypeptides and nucleic acids useful for  
XX PT vaccinating against and diagnosing infections, especially useful for  
XX PT treating acne vulgaris -  
XX PS Example 1; SEQ ID No 27794; 1069pp; English.

Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic polypeptides. The proteins and their associated DNA sequences are used in the treatment, prevention and diagnosis of medical conditions caused by P. acnes. The disorders include SAPHO syndrome (synovitis, acne, pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis. P. acnes is also involved in infections of bone, joints and the central nervous system, however it is particularly involved in the inflammatory lesions associated with acne vulgaris. A method for detecting the presence or absence of P. acnes in a patient comprises contacting a sample with a binding agent that binds to the proteins of the invention and determining the amount of bound protein in the sample. The polypeptides may be used as antigens in the production of antibodies specific for P. acnes proteins. These antibodies can be used to downregulate expression and activity of P. acnes polypeptides and therefore treat P. acnes infections. The antibodies may also be used as diagnostic agents for determining P. acnes presence, for example, by enzyme linked immunosorbent assay (ELISA).  
Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 119 AA;

QY 1 WRQTRKD 7  
||| |  
DB 8 WRQVRD 14

RESULT 7  
AAU50934  
ID AAU50934 standard; Protein; 176 AA.  
XX AC  
XX AAU50934;  
XX DT 27-FEB-2002 (first entry)  
XX DE Propionibacterium acnes immunogenic protein #11830.  
XX KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
XX KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
XX KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
XX KW dermatological; osteopathic; neuroprotectant.  
XX OS Propionibacterium acnes.  
XX PN WO200181581-A2.  
XX PD 01-NOV-2001.  
XX PF 20-APR-2001; 2001WO-US12865.  
XX PR 21-APR-2000; 2000US-199047P.  
XX PR 02-JUN-2000; 2000US-208841P.  
XX PR 07-JUL-2000; 2000US-216747P.  
XX PA (CORI-) CORIXA CORP.  
XX PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
XX PI L'maisonneuve J, Zhang Y, Jen S, Carter D;  
XX WPI; 2001-616774/71.  
XX DR N-PSDB; AAS59549.  
XX PT Propionibacterium acnes polypeptides and nucleic acids useful for  
XX PT vaccinating against and diagnosing infections, especially useful for  
XX PT treating acne vulgaris -  
XX PS Example 1; SEQ ID No 12129; 1069pp; English.

Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic polypeptides. The proteins and their associated DNA sequences are used in the treatment, prevention and diagnosis of medical conditions caused by P. acnes. The disorders include SAPHO syndrome (synovitis, acne, pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis. P. acnes is also involved in infections of bone, joints and the central nervous system, however it is particularly involved in the inflammatory lesions associated with acne vulgaris. A method for detecting the presence or absence of P. acnes in a patient comprises contacting a sample with a binding agent that binds to the proteins of the invention and determining the amount of bound protein in the sample. The polypeptides may be used as antigens in the production of antibodies specific for P. acnes proteins. These antibodies can be used to downregulate expression and activity of P. acnes polypeptides and therefore treat P. acnes infections. The antibodies may also be used as diagnostic agents for determining P. acnes presence, for example, by enzyme linked immunosorbent assay (ELISA).  
Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 176 AA;

Query Match 81.0%; Score 34; DB 22; Length 176;  
 Best Local Similarity 71.4%; Pred. No. 78;  
 Matches 5; Conservative 1; Mismatches 0; Gaps 0;

QY 1 WRQTRKD 7  
 |||:|:|  
 Db 163 WRQARRD 169

RESULT 8  
 ABU11732  
 ID ABU11732 standard; Protein; 398 AA.  
 XX  
 AC ABU11732;  
 XX  
 DT 13-FEB-2003 (first entry)  
 XX  
 DE Human MDDT polypeptide SEQ ID 679.  
 XX  
 KW MDDT; human; disease detection and treatment molecule polypeptide;  
 KW anti-inflammatory; immunosuppressive; osteopathic; cytostatic; anti-HIV;  
 KW haemostatic; nephrotropic; antianaemic; antipsoriatic; hepatotropic;  
 KW gene therapy; protein replacement therapy; cell proliferative disorder;  
 KW cancer; adenocarcinoma; leukaemia; lymphoma; myeloma; sarcoma;  
 KW anaemia; Crohn's disease; acquired immunodeficiency syndrome; AIDS;  
 KW Goodpasture's syndrome; inflammation; osteoporosis; thrombocytopaenia;  
 KW psoriasis; hepatitis.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200279449-A2.  
 XX  
 PD 10-OCT-2002.  
 XX  
 PF 27-MAR-2002; 2002WO-US09944.  
 XX  
 PR 28-MAR-2001; 2001US-279619P.  
 PR 29-MAR-2001; 2001US-280067P.  
 PR 16-MAY-2001; 2001US-291280P.  
 PR 17-MAY-2001; 2001US-291829P.  
 PR 19-JUN-2001; 2001US-291849P.  
 PR 20-JUN-2001; 2001US-299428P.  
 PR 20-JUN-2001; 2001US-299776P.  
 PR 20-JUN-2001; 2001US-300001P.  
 XX  
 PA (INCY-) INCYTE GENOMICS INC.  
 XX  
 PI Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;  
 PI Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Amshey SR;  
 PI Daugherty SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstin EH;  
 PI Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;  
 PI Flores V, Marwaha R, Lo A, Lan RY, Urashka ME;  
 XX  
 DR WPI; 2003-058431/05.  
 DR N-PSDB; ABX34722.  
 XX  
 PT New purified disease detection and treatment molecule proteins and  
 PT polynucleotides, useful for diagnosing, treating or preventing cancers  
 PT (e.g. leukemia or sarcoma), anemia, Crohn's disease, AIDS, osteoporosis  
 PT or hepatitis -

Query Match 81.0%; Score 34; DB 24; Length 398;  
 Best Local Similarity 71.4%; Pred. No. 1.7e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 WRQTRKD 7  
 |||:|:|  
 Db 74 WRKTRFD 80

RESULT 9  
 ABU11557  
 ID ABU11557 standard; Protein; 493 AA.  
 XX  
 AC ABU11557;  
 XX  
 DT 12-FEB-2003 (first entry)  
 XX  
 DE Human MDDT polypeptide SEQ ID 504.  
 XX  
 KW MDDT; human; disease detection and treatment molecule polypeptide;  
 KW anti-inflammatory; immunosuppressive; osteopathic; cytostatic; anti-HIV;  
 KW haemostatic; nephrotropic; antianaemic; antipsoriatic; hepatotropic;  
 KW gene therapy; protein replacement therapy; cell proliferative disorder;  
 KW cancer; adenocarcinoma; leukaemia; lymphoma; melanoma; sarcoma;  
 KW anaemia; Crohn's disease; acquired immunodeficiency syndrome; AIDS;  
 KW Goodpasture's syndrome; inflammation; osteoporosis; thrombocytopaenia;  
 KW psoriasis; hepatitis.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200279449-A2.  
 XX  
 PD 10-OCT-2002.  
 XX  
 PF 27-MAR-2002; 2002WO-US09944.  
 XX  
 PR 28-MAR-2001; 2001US-279619P.  
 PR 29-MAR-2001; 2001US-280067P.  
 PR 16-MAY-2001; 2001US-291280P.  
 PR 17-MAY-2001; 2001US-291829P.  
 PR 19-JUN-2001; 2001US-291849P.  
 PR 20-JUN-2001; 2001US-299428P.  
 PR 20-JUN-2001; 2001US-299776P.  
 PR 20-JUN-2001; 2001US-300001P.  
 XX  
 PA (INCY-) INCYTE GENOMICS INC.  
 XX  
 PI Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;  
 PI Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Amshey SR;  
 PI Daugherty SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstin EH;  
 PI Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;  
 PI Flores V, Marwaha R, Lo A, Lan RY, Urashka ME;  
 XX  
 DR WPI; 2003-058431/05.  
 DR N-PSDB; ABX34547.  
 XX  
 PT New purified disease detection and treatment molecule proteins and  
 PT polynucleotides, useful for diagnosing, treating or preventing cancers  
 PT (e.g. leukemia or sarcoma), anemia, Crohn's disease, AIDS, osteoporosis  
 PT or hepatitis -

Query Match 81.0%; Score 34; DB 22; Length 176;  
 Best Local Similarity 71.4%; Pred. No. 78;  
 Matches 5; Conservative 1; Mismatches 0; Gaps 0;

QY 1 WRQTRKD 7  
 |||:|:|  
 Db 163 WRQARRD 169

RESULT 8  
 ABU11732  
 ID ABU11732 standard; Protein; 398 AA.  
 XX  
 AC ABU11732;  
 XX  
 DT 13-FEB-2003 (first entry)  
 XX  
 DE Human MDDT polypeptide SEQ ID 679.  
 XX  
 KW MDDT; human; disease detection and treatment molecule polypeptide;  
 KW anti-inflammatory; immunosuppressive; osteopathic; cytostatic; anti-HIV;  
 KW haemostatic; nephrotropic; antianaemic; antipsoriatic; hepatotropic;  
 KW gene therapy; protein replacement therapy; cell proliferative disorder;  
 KW cancer; adenocarcinoma; leukaemia; lymphoma; myeloma; sarcoma;  
 KW anaemia; Crohn's disease; acquired immunodeficiency syndrome; AIDS;  
 KW Goodpasture's syndrome; inflammation; osteoporosis; thrombocytopaenia;  
 KW psoriasis; hepatitis.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200279449-A2.  
 XX  
 PD 10-OCT-2002.  
 XX  
 PF 27-MAR-2002; 2002WO-US09944.  
 XX  
 PR 28-MAR-2001; 2001US-279619P.  
 PR 29-MAR-2001; 2001US-280067P.  
 PR 16-MAY-2001; 2001US-291280P.  
 PR 17-MAY-2001; 2001US-291829P.  
 PR 19-JUN-2001; 2001US-291849P.  
 PR 20-JUN-2001; 2001US-299428P.  
 PR 20-JUN-2001; 2001US-299776P.  
 PR 20-JUN-2001; 2001US-300001P.  
 XX  
 PA (INCY-) INCYTE GENOMICS INC.  
 XX  
 PI Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;  
 PI Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Amshey SR;  
 PI Daugherty SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstin EH;  
 PI Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;  
 PI Flores V, Marwaha R, Lo A, Lan RY, Urashka ME;  
 XX  
 DR WPI; 2003-058431/05.  
 DR N-PSDB; ABX34722.  
 XX  
 PT New purified disease detection and treatment molecule proteins and  
 PT polynucleotides, useful for diagnosing, treating or preventing cancers  
 PT (e.g. leukemia or sarcoma), anemia, Crohn's disease, AIDS, osteoporosis  
 PT or hepatitis -

Query Match 81.0%; Score 34; DB 24; Length 398;  
 Best Local Similarity 71.4%; Pred. No. 1.7e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 WRQTRKD 7  
 |||:|:|  
 Db 74 WRKTRFD 80

RESULT 9  
 ABU11557  
 ID ABU11557 standard; Protein; 493 AA.  
 XX  
 AC ABU11557;  
 XX  
 DT 12-FEB-2003 (first entry)  
 XX  
 DE Human MDDT polypeptide SEQ ID 504.  
 XX  
 KW MDDT; human; disease detection and treatment molecule polypeptide;  
 KW anti-inflammatory; immunosuppressive; osteopathic; cytostatic; anti-HIV;  
 KW haemostatic; nephrotropic; antianaemic; antipsoriatic; hepatotropic;  
 KW gene therapy; protein replacement therapy; cell proliferative disorder;  
 KW cancer; adenocarcinoma; leukaemia; lymphoma; melanoma; sarcoma;  
 KW anaemia; Crohn's disease; acquired immunodeficiency syndrome; AIDS;  
 KW Goodpasture's syndrome; inflammation; osteoporosis; thrombocytopaenia;  
 KW psoriasis; hepatitis.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200279449-A2.  
 XX  
 PD 10-OCT-2002.  
 XX  
 PF 27-MAR-2002; 2002WO-US09944.  
 XX  
 PR 28-MAR-2001; 2001US-279619P.  
 PR 29-MAR-2001; 2001US-280067P.  
 PR 16-MAY-2001; 2001US-291280P.  
 PR 17-MAY-2001; 2001US-291829P.  
 PR 19-JUN-2001; 2001US-291849P.  
 PR 20-JUN-2001; 2001US-299428P.  
 PR 20-JUN-2001; 2001US-299776P.  
 PR 20-JUN-2001; 2001US-300001P.  
 XX  
 PA (INCY-) INCYTE GENOMICS INC.  
 XX  
 PI Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;  
 PI Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Amshey SR;  
 PI Daugherty SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstin EH;  
 PI Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;  
 PI Flores V, Marwaha R, Lo A, Lan RY, Urashka ME;  
 XX  
 DR WPI; 2003-058431/05.  
 DR N-PSDB; ABX34547.  
 XX  
 PT New purified disease detection and treatment molecule proteins and  
 PT polynucleotides, useful for diagnosing, treating or preventing cancers  
 PT (e.g. leukemia or sarcoma), anemia, Crohn's disease, AIDS, osteoporosis  
 PT or hepatitis -

XX PS Claim 27; SEQ ID NO 504; 339pp + Sequence Listing; English.

XX CC This invention describes a novel disease detection and treatment molecule

XX CC polypeptide (MDPT) which has anti-inflammatory, immunosuppressive,

XX CC osteopathic, cytostatic, anti-HIV, haemostatic, nephrotropic,

XX CC antianemic, antipsoriatic and hepatotropic activity. The polynucleotides

XX CC and the polypeptides of the invention can be used for gene therapy,

XX CC protein replacement therapy and are useful for treating a variety of

XX CC diseases or conditions. These polypeptides or polynucleotides are

XX CC particularly useful for diagnosing, treating or preventing cell

XX CC proliferative disorders (e.g. cancers including adenocarcinoma,

XX CC leukaemia, lymphoma, melanoma, myeloma or sarcoma), anaemia, Crohn's

XX CC disease, acquired immunodeficiency syndrome (AIDS), Goodpasture's

XX CC syndromes, inflammation, osteoporosis, thrombocytopenia, psoriasis or

XX CC hepatitis. ABU11450-ABU11845 represent the MDPT polynucleotides encoded

XX CC by ABU11450-ABU11845, described in the disclosure of the invention.

XX CC NOTE: The sequence data for this patent did not form part of the printed

XX CC specification, but was obtained in electronic format from WIPO at

XX CC ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 493 AA;

Query Match 81.0%; Score 34; DB 24; Length 493;

Best Local Similarity 71.4%; Pred. No. 2.1e+02;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 WRQTRKD 7

DB 169 WRKTRED 175

RESULT 10

AAU98880

ID AAU98880 standard; Protein; 563 AA.

AC AAU98880;

XX DT 23-AUG-2002 (first entry)

XX DE Human patched-like protein #3.

XX KW Patched-like; human; cytostatic; antidiabetic; cardiac;

XX KW neuroprotective; cancer; diabetes; cardiovascular disease;

XX KW nervous system disorder.

XX OS Homo sapiens.

XX PN WO200236613-A2.

XX PD 10-MAY-2002.

XX PF 05-NOV-2001; 2001WO-EP12778.

XX PR 06-NOV-2000; 2000US-245564P.

XX PR 06-NOV-2000; 2000US-245565P.

XX PR 06-NOV-2000; 2000US-245572P.

XX PA (FARB ) BAYER AG.

XX PI Smolyar A;

XX DR WPI; 2002-490000/52.

XX DR N-PSDB; ASK86129.

XX CC New human patched-like protein polypeptide for identifying modulating

XX PT agents useful in treating diseases e.g. cancer, diabetes,

XX PT cardiovascular diseases and central nervous system disorders -

XX CC Claim 1; Fig 10; 137pp; English.

XX PS This invention relates to the DNA and protein sequences of purified

XX CC human patched-like proteins. The nucleic acid sequences of the

CC invention are useful for making probes for detecting expression of the

CC nucleic acid in a biological sample. The protein and DNA sequences are

CC also useful for screening agents which modulate the activity of the

CC patched-like protein. A new composition or reagent containing the

CC sequences of the invention may be useful for modulating the activity of

CC the protein in a disease thereby treating a disease associated with

CC expression of patched-like protein e.g. cancer, diabetes, cardiovascular

CC diseases, and peripheral and central nervous system disorders.

CC The invention also discloses methods for detecting a coding sequence for

CC nucleic acid of the invention, for detecting the sequences in a

CC biological sample, and reducing the activity of the human protein.

CC Fusion proteins comprising the patched-like protein of the invention are

CC useful for generating antibodies and for use in various assay systems,

CC and can be used as a bait protein in a two-hybrid assay or three-hybrid

CC assay. The present sequence represents a DNA sequence encoding a

CC human patched-like protein of the invention.

XX SQ Sequence 563 AA;

Query Match 81.0%; Score 34; DB 23; Length 563;

Best Local Similarity 71.4%; Pred. No. 2.3e+02;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 WRQTRKD 7

DB 26 WRKTRED 32

RESULT 11

AAW35283

ID AAW35283 standard; Protein; 381 AA.

XX AC AAW35283;

XX DT 17-FEB-1998 (first entry)

XX DE Human acid sphingomyelinase mutant fsp330.

XX KW Prenatal diagnosis; Type A; Type B; Niemann-Pick disease;

XX KW identification; potential genetic transmitter; detection;

XX KW recessive mutation; acid sphingomyelinase; Ashkenazi Jew;

XX KW human; treatment; mutant frame shift Pro330.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX FT Region 86..88

XX FT /note= "potential N-glycosylation site"

XX FT Region 175..177

XX FT /note= "potential N-glycosylation site"

XX PN US5686240-A.

XX PD 11-NOV-1997.

XX PF 27-MAY-1994; 94US-0250740.

XX PR 27-MAY-1994; 94US-0250740.

XX PR 03-MAY-1991; 91US-0695572.

XX PA (MOUN ) MOUNT SINAI SCHOOL MEDICINE.

XX PI Desnick RJ, Schuchman EH;

XX DR WPI; 1997-558133/51.

XX DR N-PSDB; AAT95068.

XX CC Diagnosing Type A or B Niemann-Pick disease - by detecting recessive

XX PT mutation in acid sphingomyelinase gene

XX PS Disclosure; Column -; 58pp; English.

XX CC Diagnosing Type A or B Niemann-Pick disease (NPD), or identifying a

CC person as having the potential to genetically transmit Type A or B  
 CC NPD, comprises detecting a recessive mutation in the acid  
 CC sphingomyelinase (ASM) gene, which results in an alteration of at  
 CC least 1 amino acid in the ASM amino acid sequence. The method is  
 CC especially useful for prenatal diagnosis in Ashkenazi Jewish  
 CC populations. The mutation is A9G496Leu, deltaA9G608 Leu302Pro or  
 CC the fsp330 mutation described by the present sequence, where fsp330  
 CC is a frame shift mutation comprising a cytosine deletion in ASM  
 CC codon 330. The mutations are detected by selectively amplifying  
 CC mutation containing portions of the ASM gene by PCR using primers  
 CC complementary and identical to a portion of the ASM cDNA sequence,  
 CC and sequencing the amplified DNA or subjecting it to a  
 CC hybridisation assay using mutation specific probes. The ASM type 1  
 CC sequence, or the cDNA sequence encoding it can also be used in the  
 CC treatment of NPD.  
 CC N.B. Sequence not given in the specification, but constructed using  
 CC the wild type ASM sequence given in columns 56-60.

XX Sequence 381 AA;

Query Match 78.6%; Score 33; DB 18; Length 381;  
 Best Local Similarity 71.4%; Pred. No. 2.5e+02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 WRQTRKD 7

Db 285 WHQTRQD 291

RESULT 12

AAAY92343  
 ID AAY92343 standard; Protein; 472 AA.

XX AC AAY92343;

XX DT 10-AUG-2000 (first entry)

XX Human cancer associated antigen precursor from clone NY-REN-58.

XX renal cancer; cancer associated antigen precursor; diagnosis;  
 XX cytosstatic.

XX Homo sapiens.

XX WO2000020587-A2.

XX PD 13-APR-2000.

XX PF 04-OCT-1999; 99WO-US22873.

XX PR 05-OCT-1998; 98US-0166300.

XX PR 05-OCT-1998; 98US-0166350.

XX PA (LUDW-) LUDWIG INST CANCER RES.

XX Obata Y, Gout I, Tureci O, Sahin U, Pfreundschuh M, Scanlan MJ;

XX Stockert E, Chen Y, Old LJ, Jager E, Knuth A;

XX WPI; 2000-303774/26.

XX DR N-PSDB; AAA03315.

XX Preventing, diagnosing and/or treating disorders associated with

XX abnormal expression of human cancer associated antigens

XX Example 1; Page 89-90; 121pp; English.

XX AAY92338-47 are encoded by novel genes isolated by SEREX screening from  
 CC a renal cancer cell line 1973/10.4. They are cancer associated antigen  
 CC precursors. These gene products are useful in methods for preventing,  
 CC diagnosing and/or treating disorders, especially cancer, associated with  
 CC abnormal expression of human cancer associated antigens. The method  
 CC comprises contacting a sample from a subject with an agent that  
 CC specifically binds to the nucleic acid molecule or expression product

CC (or fragment) complexed with a human leukocyte antigen (HLA) molecule  
 CC and determining the interaction between the agent and the nucleic acid  
 CC molecule or the expression product as a determination of the disorder.

XX Sequence 472 AA;

Query Match 78.6%; Score 33; DB 21; Length 472;

Best Local Similarity 71.4%; Pred. No. 3.1e+02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 WRQTRKD 7

Db 289 WRQSEKD 295

RESULT 13

AAR30644  
 ID AAR30644 standard; Protein; 628 AA.

XX AC AAR30644;

XX DT 25-MAR-2003 (updated)

XX DT 06-MAY-1993 (first entry)

XX DE Deltar608 ASM.

XX Acid sphingomyelinase; ASM; type; 1; 2; PCR; primer; amplify; cryptic;  
 KW polymerase chain reaction; splice site; mutation; R496L; Deltar608;  
 KW L302; Neimann-Pick disease; NPD; jewish community.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Misc-difference 607..608

FT /label= Deltar608

XX PN EF520843-A2.

XX PD 30-DEC-1992.

XX PF 30-APR-1992; 92EP-0401241.

XX PR 03-MAY-1991; 91US-0695472.

XX PA (MOUN ) MOUNT SINAI MEDICAL CENT.

XX Desnick RJ, Schuchman EH;

XX WPI; 1993-001632/01.

XX DR N-PSDB; AAQ33394.

XX Pure and recombinant acid sphingomyelinase and its nucleic acid -  
 PT for treatment and diagnosis of Niemann-Pick disease

XX Claims 5 and 20; Fig 3; 50pp; English.

XX This sequence represents the acid sphingomyelinase (ASM) mutation  
 CC delta R608. This ASM gene sequence was isolated by PCR using primers  
 CC constructed from common exonic sequences flanking the type 1 and 2  
 CC specific sequences. Certain mutations in the ASM gene ie. R496L,  
 CC Deltar608 and L302 have been found to correlate with Neimann-Pick  
 CC disease (NPD). See also AAQ33390-423.  
 CC (Updated on 25-MAR-2003 to correct PN field.)

XX SQ Sequence 628 AA;

Query Match 78.6%; Score 33; DB 14; Length 628;

Best Local Similarity 71.4%; Pred. No. 4e+02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 WRQTRKD 7

Db 285 WHQTRQD 291



RESULT 14  
AAW35262  
ID AAW35262 standard; Protein; 628 AA.  
XX AC AAW35262;  
XX 17-FEB-1998 (first entry)  
XX DE Human acid sphingomyelinase mutant deltaArg608.  
XX Prenatal diagnosis; Type A; Type B; Niemann-Pick disease;  
KW identification; potential genetic transmitter; detection;  
KW recessive mutation; acid sphingomyelinase; Ashkenazi Jew;  
XX human; treatment; mutant deltaArg608.  
XX OS Homo sapiens.  
XX FH Key  
FT Region  
FT /note= "potential N-glycosylation site"  
FT Region  
FT /note= "potential N-glycosylation site"  
FT Region  
FT /note= "potential N-glycosylation site"  
FT Region  
FT /note= "potential N-glycosylation site"  
FT Region  
FT /note= "potential N-glycosylation site"  
FT Region  
FT /note= "potential N-glycosylation site"  
FT Region  
FT /note= "potential N-glycosylation site"  
XX US5686240-A.  
XX 11-NOV-1997.  
XX 27-MAY-1994; 94US-0250740.  
XX 27-MAY-1994; 94US-0250740.  
PR 03-MAY-1991; 91US-0695572.  
XX (MOUN ) MOUNT SINAI SCHOOL MEDICINE.  
XX Desnick RJ, Schuchman EH;  
XX WPI; 1997-558133/51.  
DR N-PSDB; AAT95066.  
XX Diagnosing Type A or B Niemann-Pick disease - by detecting recessive  
PT mutation in acid sphingomyelinase gene  
XX Disclosure; Column -: 58pp; English.  
XX Diagnosing Type A or B Niemann-Pick disease (NPD), or identifying a  
CC person as having the potential to genetically transmit Type A or B  
CC NPD, comprises detecting a recessive mutation in the acid  
CC sphingomyelinase (ASM) gene, which results in an alteration of at  
CC least 1 amino acid in the ASM amino acid sequence. The method is  
CC especially useful for prenatal diagnosis in Ashkenazi Jewish  
CC populations. The mutation is Arg496Leu, the deltaArg608 mutation  
CC described by the present sequence, Leu302Pro or Isp330, where  
CC fsp330 is a frame shift mutation comprising a cytosine deletion in  
CC ASM codon 330. The mutations are detected by selectively amplifying  
CC mutation containing portions of the ASM gene by PCR using primers  
CC complementary and identical to a portion of the ASM cDNA sequence,  
CC and sequencing the amplified DNA or subjecting it to a  
CC hybridisation assay using mutation specific probes. The ASM type 1  
CC sequence, or the cDNA sequence encoding it can also be used in the  
CC treatment of NPD.  
CC N.B. Sequence not given in the specification, but constructed using  
CC the wild type ASM sequence given in columns 56-60.

SQ Sequence 628 AA;  
Query Match 78.6%; Score 33; DB 18; Length 628;  
Best Local Similarity 71.4%; Pred. No. 4e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 WRQTRKD 7  
DB 285 WHQTRQD 291  
RESULT 15  
AAR30642  
ID AAR30642 standard; Protein; 629 AA.  
XX AC AAR30642;  
XX 25-MAR-2003 (updated)  
DT 06-MAY-1993 (first entry)  
XX ASM protein.  
XX Acid sphingomyelinase; ASM; pASM-1FL; recombinant DNA; R496L; NPD;  
KW deltaR608; deltaL302; Neimann-Pick disease; Jewish community.  
XX OS Homo sapiens.  
XX FH Key  
FT Peptide  
FT /note= "Type 1 sequence"  
FT Peptide  
FT /note= "Tryptic peptide, T1"  
FT Peptide  
FT /note= "Tryptic peptide, T2"  
FT Peptide  
FT /note= "Tryptic peptide, T3"  
FT Peptide  
FT /note= "Tryptic peptide, T4"  
FT Peptide  
FT /note= "Tryptic peptide, T5"  
FT Peptide  
FT /note= "Tryptic peptide, T6"  
FT Peptide  
FT /note= "Tryptic peptide, T7"  
FT Peptide  
FT /note= "Tryptic peptide, T8"  
FT Peptide  
FT /note= "Tryptic peptide, T9"  
FT Peptide  
FT /note= "Tryptic peptide, T10"  
FT Peptide  
FT /note= "Tryptic peptide, T11"  
FT Peptide  
FT /note= "Tryptic peptide, T12"  
FT Modified-site  
FT /note= "Potential N-glycosylation site"  
FT Modified-site  
FT /note= "Potential N-glycosylation site"  
FT Modified-site  
FT /note= "Potential N-glycosylation site"  
FT Modified-site  
FT /note= "Potential N-glycosylation site"  
FT Modified-site  
FT /note= "Potential N-glycosylation site"  
FT Modified-site  
FT /note= "Potential N-glycosylation site"  
FT Misc-difference  
FT /note= "Amino acid varies from that found in  
FT fibroblast cDNA's pASM-1 and pASM-2"  
FT Misc-difference  
FT /note= "Amino acid varies from that found in  
FT fibroblast cDNA's pASM-1 and pASM-2"  
XX

Thu Feb 12 09:07:26 2004

```
PN EP520843-A2.
XX
XX
XX 30-DEC-1992.
XX
XX 30-APR-1992; 92EP-0401241.
XX
XX 03-MAY-1991; 91US-0695472.
XX
XX (MOUN ) MOUNT SINAI MEDICAL CENT.
XX
XX Desnick RJ, Schuchman EH;
XX
XX WPI; 1993-001632/01.
XX N-PSDB; AAQ33390.
XX
XX Pure and recombinant acid sphingomyelinase and its nucleic acid -
XX for treatment and diagnosis of Niemann-Pick disease
XX
XX Claim 14; Fig 3; 50pp; English.
XX
XX This sequence represents functional acid sphingomyelinase (ASM) and
XX was encoded by plasmid pASM-1FL. The nucleotide sequence encoding
XX this protein may be used to generate recombinant DNA molecules that
XX direct expression of the enzyme product. Certain mutations in the
XX ASM gene ie. R496L, deltaR608 and I302 have been found to correlate
XX with Niemann-Pick disease (NPD). See also AAQ33391-423.
XX (Updated on 25-MAR-2003 to correct PN field.)
XX
XX Sequence 629 AA;
SQ
Query Match 78.6%; Score 33; DB 14; Length 629;
Best Local Similarity 71.4%; Pred. NO. 4e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
OY 1 WQTRKD 7
Db 285 WQTRKD 291
Search completed: February 11, 2004, 17:02:44
Job time : 34.25 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 16:52:34 ; Search time 24.5833 Seconds  
(without alignments)  
73.479 Million cell updates/sec

Title: US-09-901-187C-2

Perfect score: 41

Sequence: 1 HYAKNPI 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL.23.\*

1: sp\_archaea.\*

2: sp\_bacteria.\*

3: sp\_fungi.\*

4: sp\_human.\*

5: sp\_invertebrate.\*

6: sp\_mammal.\*

7: sp\_mhc.\*

8: sp\_organelle.\*

9: sp\_phase.\*

10: sp\_plant.\*

11: sp\_rodent.\*

12: sp\_virus.\*

13: sp\_vertebrate.\*

14: sp\_unclassified.\*

15: sp\_rvirus.\*

16: sp\_bacteriap.\*

17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37	90.2	97	16	Q9CEQ1
2	37	90.2	516	5	Q95290
3	37	90.2	1675	5	Q9VLR1
4	36	87.8	127	8	Q99881
5	36	87.8	209	8	Q922M0
6	36	87.8	252	8	Q954U2
7	36	87.8	253	8	Q9MM56
8	36	87.8	253	8	Q9MD65
9	36	87.8	253	8	Q9MM57
10	36	87.8	253	8	Q9MM51
11	36	87.8	253	8	Q9T6Q6
12	36	87.8	253	8	Q9MM58
13	36	87.8	253	8	Q9MM54
14	36	87.8	253	8	Q9MM59
15	36	87.8	253	8	Q9MM55
16	36	87.8	253	8	Q9MM53

17	36	87.8	257	8	Q00809
18	36	87.8	258	8	Q99029
19	36	87.8	259	8	Q948B5
20	36	87.8	260	8	Q37758
21	36	87.8	306	8	Q99032
22	36	87.8	317	8	Q9MFA2
23	36	87.8	361	8	Q31737
24	35	85.4	420	10	Q8S8H8
25	35	85.4	1937	12	Q91BP7
26	35	85.4	2860	5	Q81640
27	34	82.9	342	1	Q9UXN9
28	34	82.9	346	17	Q8ETC7
29	34	82.9	1134	10	Q9FH48
30	34	82.9	2064	12	Q8B6R7
31	34	82.9	2919	12	Q85431
32	33	80.5	160	2	Q9R9C7
33	33	80.5	176	16	Q9WZC4
34	33	80.5	197	2	Q44568
35	33	80.5	200	16	Q25706
36	33	80.5	237	15	Q8JAH0
37	33	80.5	269	11	Q9CT18
38	33	80.5	383	16	Q914H5
39	33	80.5	386	16	Q8XZK5
40	33	80.5	460	4	Q96HD7
41	33	80.5	465	10	Q9SDZ9
42	33	80.5	485	10	Q9ZS50
43	33	80.5	522	16	Q8IF92
44	33	80.5	593	13	Q90687
45	33	80.5	595	13	Q92124

ALIGNMENTS

RESULT 1

Q9CEQ1 PRELIMINARY; PRT; 97 AA.

AC Q9CEQ1; 01-JUN-2001 (Tremblrel. 17, Created)  
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)  
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)  
DE Unknown protein.  
GN YSED OR L11785.  
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.  
OX NCBI\_TaxID=1360;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=IL1403;  
RX MEDLINE=21235186; PubMed=11337471;  
RA Bolotin A., Wincker P., Manger S., Jaillon O., Malarne K.,  
RA Weissenbach J., Ehrlich S.D., Sorokin A.  
RT "The complete genome sequence of the lactic acid bacterium Lactococcus  
lactis ssp. lactis IL1403".  
RL Genome Res. 11:731-753(2001).  
DR EMBL; AE006408; AAK05883.1; -.  
KW Complete proteome.  
SQ SEQUENCE 97 AA; 11573 MW; 7FF30A958C52F1E3 CRC64;

Query Match 90.2%; Score 37; DB 16; Length 97;  
Best Local Similarity 85.7%; Pred No. 3.5;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HYAKNPI 7

Db 26 HYGNPI 32

RESULT 2

Q95290 PRELIMINARY; PRT; 516 AA.

ID Q95290

AC Q95290;

DT 01-DEC-2001 (Tremblrel. 19, Created)

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DE Hypochemical 57.8 kDa protein.  
 GN LT 05.  
 OS Leishmania major.  
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
 OX NCBI\_TaxID=5664;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Friedlin;  
 RA Ivens A.C., Murphy L., Quail M., Harris D., Oliver K., Lawson D.,  
 RA Rajandream M.A., Barrall B.G.;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Friedlin;  
 RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,  
 RA Smith D.F.;  
 RT "A physical map of the Leishmania major Friedlin genome."  
 RL Genome Res. 8:135-145(1998).  
 DR EMBL; AL389894; CAC44913.1;  
 KW Hypochemical protein.  
 SQ SEQUENCE 516 AA; 57779 MW; DDB9A9615AB4865 CRC64;  
 Query Match 90.2%; Score 37; DB 5; Length 516;  
 Best Local Similarity 100.0%; Pred. No. 20;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 HYAKNP 6  
 Db 83 HYAKNP 88  
 RESULT 3  
 Q9VLT1 PRELIMINARY; PRT; 1675 AA.  
 AC Q9VLT1; Q9VLT2;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE CG31756 protein.  
 GN CG8683 OR CGI4276.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer V.G., Champagne M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasly E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.N., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delella A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foeller C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glöck A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Keichum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

LAsko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarri C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S. Yao Q.A.,  
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster."  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,  
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,  
 RA Carlson J.W., Center A., Champagne M., Davenport L.B., Dietz S.M.,  
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,  
 RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,  
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
 RA Ibegwan C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
 RA Pacleb J., Parag V., Park S., Patel S., Pfeiffer B.,  
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,  
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
 RT "Sequencing of Drosophila melanogaster genome."  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,  
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,  
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
 RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,  
 RA Searle S.M.J., Smith E., Shu S., Smutniak P., Whitfield E.,  
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
 RT "Annotation of Drosophila melanogaster genome."  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA Flybase;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AE003620; AAF52603.3;  
 DR FlyBase; FBgn0031985; CG8683.  
 SQ SEQUENCE 1675 AA; 185905 MW; EA5A3DD1943FF96E CRC64;  
 Query Match 90.2%; Score 37; DB 5; Length 1675;  
 Best Local Similarity 100.0%; Pred. No. 70;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 HYAKNP 6  
 Db 155 HYAKNP 160  
 RESULT 4  
 Q99881 PRELIMINARY; PRT; 127 AA.  
 ID Q99881  
 AC Q99881  
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)

DE Cytochrome oxidase subunit 2 (Fragment).  
GN COXII.  
OS Solanum tuberosum (Potato).  
OG Mitochondrion.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Asteridae; lamids; Solanales; Solanaceae; Solanum.  
OX NCBI\_TaxID=4113;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Loessl A., Adler N., Horn R., Frei U., Wenzel G.;  
RT "Chondriome type characterization of potato at alpha, beta, gamma,  
RT delta, epsilon and novel plastid-mitochondrial configurations in  
RT somatic hybrids.";  
RL Theor. Appl. Genet. 0:0-0(1999).  
DR EMBL; AF096321; AAD03043.1; -.  
DR InterPro; IPR002429; Cyt\_c\_ox\_2.  
DR Pfam; PF00116; COX2; 1.  
DR Pfam; PF02790; COX2\_TM; 1.  
DR PRINTS; PRO1166; CYCOXIDASEII.  
DR Oxidoreductase; Transmembrane; Mitochondrion.  
KW Oxidoreductase; Transmembrane; Mitochondrion.  
FT NON\_TER 127  
SQ SEQUENCE 127 AA; 14582 MW; F7FDC736B618DEB1 CRC64;  
Query Match 87.8%; Score 36; DB 8; Length 127;  
Best Local Similarity 85.7%; Pred. No. 7.6; 1; Indels 0; Gaps 0;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 HYAKNPI 7  
DB 65 HYKKNPI 71  
RESULT 5  
ID Q9Z2M0 PRELIMINARY; PRT; 209 AA.  
AC Q9Z2M0;  
DT 01-MAY-1999 (TRENBLrel. 10, Created)  
DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)  
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
DE Cytochrome c oxidase subunit II (EC 1.9.3.1) (Cytochrome c oxidase  
DE polypeptide II) (Fragment).  
GN COXII.  
OS Magnolia pyramidalata.  
OG Mitochondrion.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Magnoliaceae; Magnoliaceae;  
OX NCBI\_TaxID=44759;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Jones D.V., Thien L.B., Latimer S., Hurley D.L.;  
RT "Loss of the coxII intron in Magnolia pyramidalata and M. tripetala  
RT (Magnoliaceae).";  
RL Submitted (DSC-1995) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY  
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-  
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2  
CC TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER  
CC A CENTER TO THE BIMETALLIC CENTER OF THE CATALYTIC SUBUNIT 1 (BY  
CC SIMILARITY).  
CC -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 4 FERRICYTOCHROME  
CC C + 2 H(2)O.  
CC -!- COFACTOR: COPPER A (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL  
CC INNER MEMBRANE (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.  
DR EMBL; U42696; AAD00020.1; -.  
DR InterPro; IPR001920; Asp/Glu\_rac.  
DR InterPro; IPR001505; Copper\_Cua.  
DR InterPro; IPR002429; Cyt\_c\_ox\_2.  
DR Pfam; PF00116; COX2; 1.  
DR PRINTS; PRO1166; CYCOXIDASEII.  
DR Oxidoreductase; Transmembrane; Mitochondrion.  
KW Oxidoreductase; Transmembrane; Mitochondrion.  
FT NON\_TER 127  
SQ SEQUENCE 127 AA; 14582 MW; F7FDC736B618DEB1 CRC64;  
Query Match 87.8%; Score 36; DB 8; Length 127;  
Best Local Similarity 85.7%; Pred. No. 7.6; 1; Indels 0; Gaps 0;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 HYAKNPI 7  
DB 65 HYKKNPI 71

DR ProDom; PD000131; Copper\_Cua; 1.  
DR PROSITE; PS00924; ASP\_GLU\_RACEMASE\_2; 1.  
KW Copper; Electron transport; Inner membrane; Membrane; Oxidoreductase;  
KW Respiratory chain; Transmembrane; Transport; Mitochondrion.  
FT NON\_TER 1  
FT NON\_TER 209  
SQ SEQUENCE 209 AA; 23828 MW; 2B7CFC58DEF3CE39 CRC64;  
Query Match 87.8%; Score 36; DB 8; Length 209;  
Best Local Similarity 85.7%; Pred. No. 13;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 HYAKNPI 7  
DB 47 HYKKNPI 53  
RESULT 6  
ID Q954U2 PRELIMINARY; PRT; 252 AA.  
AC Q954U2;  
DT 01-DEC-2001 (TRENBLrel. 19, Created)  
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)  
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
DE Truncated cytochrome c oxidase subunit 2 (EC 1.9.3.1) (Cytochrome c  
DE oxidase polypeptide II).  
GN COX2.  
OS Beta vulgaris subsp. maritima (sea beet).  
OG Mitochondrion.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Caryophyllidae; Caryophyllales; Chenopodiaceae; Beta.  
OX NCBI\_TaxID=62760;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21287185; PubMed=11389758;  
RA Ducos E., Touzet P., Boutry M.;  
RT "The male sterile G cytoplasm of wild beet displays modified  
RT mitochondrial respiratory complexes.";  
RL Plant J. 26:171-180(2001).  
CC -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY  
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-  
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2  
CC TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER  
CC A CENTER TO THE BIMETALLIC CENTER OF THE CATALYTIC SUBUNIT 1 (BY  
CC SIMILARITY).  
CC -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 4 FERRICYTOCHROME  
CC C + 2 H(2)O.  
CC -!- COFACTOR: COPPER A (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL  
CC INNER MEMBRANE (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.  
DR EMBL; AF276430; AAK82658.1; -.  
DR InterPro; IPR001505; Copper\_Cua.  
DR InterPro; IPR002429; Cyt\_c\_ox\_2.  
DR Pfam; PF00116; COX2; 1.  
DR Pfam; PF02790; COX2\_TM; 1.  
DR PRINTS; PRO1166; CYCOXIDASEII.  
DR ProDom; PD000131; Copper\_Cua; 1.  
DR PROSITE; PS00078; COX2; 1.  
DR Copper; Electron transport; Inner membrane; Membrane; Oxidoreductase;  
KW Respiratory chain; Transmembrane; Transport; Mitochondrion.  
SQ SEQUENCE 252 AA; 28724 MW; 7C23CDFD783F8C85 CRC64;  
Query Match 87.8%; Score 36; DB 8; Length 252;  
Best Local Similarity 85.7%; Pred. No. 16;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 HYAKNPI 7  
DB 65 HYKKNPI 71

```

RESULT 7
Q9MM56 PRELIMINARY; PRT; 253 AA.
AC Q9MM56; 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Cytochrome oxidase subunit 2 (EC 1.9.3.1) (Cytochrome c oxidase
DE polypeptide II).
GN COX2.
OS Terminus uncinatus.
OG Mitochondrion.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Terminus.
OC NCBI_TaxID=109223;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20040642; PubMed=10570164;
RA Adams K.L., Song K., Roessler P.G., Nugent J.M., Doyle J.L.,
RA Doyle J.J., Palmer J.D.;
RT "Intracellular gene transfer in action: dual transcription and
RT multiple silencings of nuclear and mitochondrial cox2 genes in
RT legumes";
RL Proc. Natl. Acad. Sci. U.S.A. 96:13863-13868(1999).
CC -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2
CC TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER
CC A CENTER TO THE BIMETALLIC CENTER OF THE CATALYTIC SUBUNIT 1 (BY
CC SIMILARITY).
CC -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 4 FERRICYTOCHROME
CC C + 2 H(2)O.
CC -!- COPACITOR: COPPER A (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC INNER MEMBRANE (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.
CC EMBL; AF207680; AAF43635.1; -
CC EMBL; AF208163; AAF27822.1; -
CC EMBL; AF207673; AAF43631.1; -
CC EMBL; AF207679; AAF43637.1; -
CC InterPro; IPR001505; Copper_CuA.
CC InterPro; IPR002429; Cyt_c_Ox_2.
CC Pfam; PF00116; COX2; 1.
CC Pfam; PF02790; COX2_TM; 1.
CC PRINTS; PR01166; CYCOXIDASEII.
CC PROSITE; PD000131; Copper_CuA; 1.
CC PROSITE; PS00078; COX2; 1.
KW Copper; Electron transport; Inner membrane; Membrane; Oxidoreductase;
KW Respiratory chain; Transmembrane; Transport; Mitochondrion.
SQ SEQUENCE 253 AA; 28984 MW; 43A6551D85245091 CRC64;

Query Match 87.8%; Score 36; DB 8; Length 253;
Best Local Similarity 85.7%; Pred. No. 16;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HYAKNPI 7
Db 64 HYKKNPI 70

RESULT 8
Q9MD65 PRELIMINARY; PRT; 253 AA.
AC Q9MD65; 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Cytochrome oxidase subunit 2 (EC 1.9.3.1) (Cytochrome c oxidase
DE polypeptide II).
GN COX2.
OS Calopogonium caeruleum,
OS Pseudovigna argentea,
OS Pseudemia comosa, and
OS Pachyrhizus erosus.

```

```

OG Mitochondrion.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
OC Calopogonium.
OC NCBI_TaxID=109163, 109173, 45681, 109171;
RN [1]
RP SEQUENCE FROM N.A.
RX SPECIES=C.caeruleum, P.argentea, P.comosa, and P.erosus;
RX MEDLINE=20040642; PubMed=10570164;
RA Adams K.L., Song K., Roessler P.G., Nugent J.M., Doyle J.L.,
RA Doyle J.J., Palmer J.D.;
RT "Intracellular gene transfer in action: dual transcription and
RT multiple silencings of nuclear and mitochondrial cox2 genes in
RT legumes";
RL Proc. Natl. Acad. Sci. U.S.A. 96:13863-13868(1999).
CC -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2
CC TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER
CC A CENTER TO THE BIMETALLIC CENTER OF THE CATALYTIC SUBUNIT 1 (BY
CC SIMILARITY).
CC -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 4 FERRICYTOCHROME
CC C + 2 H(2)O.
CC -!- COPACITOR: COPPER A (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC INNER MEMBRANE (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.
CC EMBL; AF207680; AAF43638.1; -
CC EMBL; AF208163; AAF27822.1; -
CC EMBL; AF207673; AAF43631.1; -
CC EMBL; AF207679; AAF43637.1; -
CC InterPro; IPR001505; Copper_CuA.
CC InterPro; IPR002429; Cyt_c_Ox_2.
CC Pfam; PF00116; COX2; 1.
CC Pfam; PF02790; COX2_TM; 1.
CC PRINTS; PR01166; CYCOXIDASEII.
CC PROSITE; PD000131; Copper_CuA; 1.
CC PROSITE; PS00078; COX2; 1.
KW Copper; Electron transport; Inner membrane; Membrane; Oxidoreductase;
KW Respiratory chain; Transmembrane; Transport; Mitochondrion.
SQ SEQUENCE 253 AA; 29061 MW; B6A4381D85219B90 CRC64;

Query Match 87.8%; Score 36; DB 8; Length 253;
Best Local Similarity 85.7%; Pred. No. 16;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HYAKNPI 7
Db 64 HYKKNPI 70

RESULT 9
Q9MM57 PRELIMINARY; PRT; 253 AA.
AC Q9MM57; 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Cytochrome oxidase subunit 2 (EC 1.9.3.1) (Cytochrome c oxidase
DE polypeptide II).
GN COX2.
OS Neotonia wightii.
OG Mitochondrion.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Neotonia.
OC NCBI_TaxID=103823;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20040642; PubMed=10570164;
RA Adams K.L., Song K., Roessler P.G., Nugent J.M., Doyle J.L.,
RA Doyle J.J., Palmer J.D.;

```

RT "Intracellular gene transfer in action: dual transcription and multiple silencings of nuclear and mitochondrial cox2 genes in legumes.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 96:13863-13868(1999).  
 RL CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-  
 CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2  
 CC TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER  
 CC A CENTER TO THE BIMETALLIC CENTER OF THE CATALYTIC SUBUNIT 1 (BY  
 CC SIMILARITY).  
 CC -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 4 FERRICYTOCHROME  
 CC C + 2 H(2)O.  
 CC -!- COFACTOR: COPPER A (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL  
 CC INNER MEMBRANE (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.  
 DR EMBL; AF207676; AAF43634.1; -  
 DR InterPro; IPR001505; Copper\_Cua.  
 DR Pfam; PF00116; COX2; 1.  
 DR PRINTS; PR00116; CYCOXIDASEII.  
 DR PRODOM; PD000131; Copper\_Cua; 1.  
 DR PROSITE; PS00078; COX2; 1.  
 DR Copper; Electron transport; Inner membrane; Membrane; Oxidoreductase;  
 KW Respiratory chain; Transmembrane; Transport; Mitochondrion.  
 SQ SEQUENCE 253 AA; 28824 MW; 0CB4CC66A9DA5A72 CRC64;  
 Query Match 87.8%; Score 36; DB 8; Length 253;  
 Best Local Similarity 85.7%; Pred. No. 16;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 HYAKNPI 7  
 DB 64 HYKKNPI 70  
 RESULT 10  
 Q9MM51 PRELIMINARY; PRT; 253 AA.  
 ID Q9MM51;  
 AC Q9MM51; (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
 DE Cytochrome oxidase subunit 2 (EC 1.9.3.1) (Cytochrome c oxidase  
 DE polypeptide II).  
 GN COX2.  
 OS Lespedeza formosa.  
 OG Mitochondrion.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Desmodieae; Lespedeza.  
 OX NCBI\_TaxID=109221;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20040642; PubMed=10570164;  
 RA Adams K.L., Song K., Roesler P.G., Nugent J.M., Doyle J.L.,  
 RA Doyle J.J., Palmer J.D.;  
 RT "Intracellular gene transfer in action: dual transcription and  
 RT multiple silencings of nuclear and mitochondrial cox2 genes in  
 RT legumes.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 96:13863-13868(1999).  
 CC -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY  
 CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-  
 CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2  
 CC TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER  
 CC A CENTER TO THE BIMETALLIC CENTER OF THE CATALYTIC SUBUNIT 1 (BY  
 CC SIMILARITY).  
 CC -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 4 FERRICYTOCHROME  
 CC C + 2 H(2)O.  
 CC -!- COFACTOR: COPPER A (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL  
 CC INNER MEMBRANE (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.  
 DR EMBL; AF207746; AAF25663.1; -  
 DR InterPro; IPR001505; Copper\_Cua.  
 DR Pfam; PF00116; COX2; 1.  
 DR PRINTS; PR00116; CYCOXIDASEII.  
 DR PRODOM; PD000131; Copper\_Cua; 1.  
 DR PROSITE; PS00078; COX2; 1.  
 DR Copper; Electron transport; Inner membrane; Membrane; Oxidoreductase;  
 KW Respiratory chain; Transmembrane; Transport; Mitochondrion.  
 SQ SEQUENCE 253 AA; 29128 MW; A194380852E313B CRC64;  
 Query Match 87.8%; Score 36; DB 8; Length 253;  
 Best Local Similarity 85.7%; Pred. No. 16;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 HYAKNPI 7  
 DB 64 HYKKNPI 70

DR EMBL; AF208162; AAF27821.1; -  
 DR InterPro; IPR001505; Copper\_Cua.  
 DR Pfam; PF00116; COX2; 1.  
 DR PRINTS; PR00116; CYCOXIDASEII.  
 DR PRODOM; PD000131; Copper\_Cua; 1.  
 DR PROSITE; PS00078; COX2; 1.  
 DR Copper; Electron transport; Inner membrane; Membrane; Oxidoreductase;  
 KW Respiratory chain; Transmembrane; Transport; Mitochondrion.  
 SQ SEQUENCE 253 AA; 29095 MW; 16A4381F789740F9 CRC64;  
 Query Match 87.8%; Score 36; DB 8; Length 253;  
 Best Local Similarity 85.7%; Pred. No. 16;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 HYAKNPI 7  
 DB 64 HYKKNPI 70  
 RESULT 11  
 Q9T6Q6 PRELIMINARY; PRT; 253 AA.  
 ID Q9T6Q6  
 AC Q9T6Q6; (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
 DE Cytochrome oxidase subunit 2 (EC 1.9.3.1) (Cytochrome c oxidase  
 DE polypeptide II).  
 GN COX2.  
 OS Cologania lemmonii.  
 OG Mitochondrion.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Cologania.  
 OX NCBI\_TaxID=109165;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20040642; PubMed=10570164;  
 RA Adams K.L., Song K., Roesler P.G., Nugent J., Doyle J.L., Doyle J.J.,  
 RA Palmer J.D.;  
 RT "Intracellular gene transfer in action: Dual transcription and  
 RT multiple silencings of nuclear and mitochondrial cox2 genes in  
 RT legumes.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 96:13863-13868(1999).  
 CC -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY  
 CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-  
 CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2  
 CC TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER  
 CC A CENTER TO THE BIMETALLIC CENTER OF THE CATALYTIC SUBUNIT 1 (BY  
 CC SIMILARITY).  
 CC -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 4 FERRICYTOCHROME  
 CC C + 2 H(2)O.  
 CC -!- COFACTOR: COPPER A (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL  
 CC INNER MEMBRANE (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.  
 DR EMBL; AF207746; AAF25663.1; -  
 DR InterPro; IPR001505; Copper\_Cua.  
 DR Pfam; PF00116; COX2; 1.  
 DR PRINTS; PR00116; CYCOXIDASEII.  
 DR PRODOM; PD000131; Copper\_Cua; 1.  
 DR PROSITE; PS00078; COX2; 1.  
 DR Copper; Electron transport; Inner membrane; Membrane; Oxidoreductase;  
 KW Respiratory chain; Transmembrane; Transport; Mitochondrion.  
 SQ SEQUENCE 253 AA; 29128 MW; A194380852E313B CRC64;  
 Query Match 87.8%; Score 36; DB 8; Length 253;  
 Best Local Similarity 85.7%; Pred. No. 16;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 HYAKNPI 7  
 DB 64 HYKKNPI 70

Thu Feb 12 09:07:37 2004

us-09-901-187c-2.rspt

QY 1 HYAKNPI 7  
 DB 64 HYKNPI 70  
 RESULT 12  
 Q9MM58 PRELIMINARY; PRT; 253 AA.  
 AC Q9MM58;  
 DT 01-OCT-2000 (TRENBLrel. 15, Created)  
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
 DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)  
 DE Cytochrome oxidase subunit 2 (EC 1.9.3.1) (Cytochrome c oxidase polypeptide II).  
 GN Neotonia wightii.  
 OS Neotonia wightii.  
 OG Mitochondrion.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Neotonia.  
 OC NCBI\_TaxID=103923;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20040642; PubMed=10570164;  
 RA Adams K.L., Song K., Roessler P.G., Nugent J.M., Doyle J.L.,  
 RA Doyle J.J., Palmer J.D.;  
 RT "Intracellular gene transfer in action: dual transcription and multiple silencings of nuclear and mitochondrial cox2 genes in legumes.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 96:13863-13868 (1999).  
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2 TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER A CENTER TO THE BIMETALLIC CENTER OF THE CATALYTIC SUBUNIT 1 (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 4 FERRICYTOCHROME C + 2 H(2)O.  
 CC -1- COFACTOR: COPPER A (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL INNER MEMBRANE (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY. EMBL; AF207675; AAF43633.1; InterPro; IPR001505; Copper\_CuA. InterPro; IPR002429; Cyt\_c\_ox\_2. Pfam; PF00116; COX2; 1. PRINTS; PR01166; CYCOXIDASEII. PRODOM; PD000131; Copper\_CuA; 1. PROSITE; PS00078; COX2; 1. Copper; Electron transport; Inner membrane; Membrane; Oxidoreductase; Respiratory chain; Transmembrane; Transport; Mitochondrion. SEQUENCE 253 AA; 29095 MW; BBE9F70ED865C52 CRC64;  
 Query Match 87.8%; Score 36; DB 8; Length 253;  
 Best Local Similarity 85.7%; Pred. No. 16;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 HYAKNPI 7  
 DB 64 HYKNPI 70  
 RESULT 13  
 Q9MM54 PRELIMINARY; PRT; 253 AA.  
 AC Q9MM54;  
 DT 01-OCT-2000 (TRENBLrel. 15, Created)  
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
 DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)  
 DE Cytochrome oxidase subunit 2 (EC 1.9.3.1) (Cytochrome c oxidase polypeptide II).  
 GN Neotonia wightii.  
 OS Neotonia wightii.  
 OG Mitochondrion.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Neotonia.  
 OC NCBI\_TaxID=103923;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20040642; PubMed=10570164;  
 RA Adams K.L., Song K., Roessler P.G., Nugent J.M., Doyle J.L.,  
 RA Doyle J.J., Palmer J.D.;  
 RT "Intracellular gene transfer in action: dual transcription and multiple silencings of nuclear and mitochondrial cox2 genes in legumes.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 96:13863-13868 (1999).  
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2 TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER A CENTER TO THE BIMETALLIC CENTER OF THE CATALYTIC SUBUNIT 1 (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 4 FERRICYTOCHROME C + 2 H(2)O.  
 CC -1- COFACTOR: COPPER A (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL INNER MEMBRANE (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY. EMBL; AF207675; AAF43633.1; InterPro; IPR001505; Copper\_CuA. InterPro; IPR002429; Cyt\_c\_ox\_2. Pfam; PF00116; COX2; 1. PRINTS; PR01166; CYCOXIDASEII. PRODOM; PD000131; Copper\_CuA; 1. PROSITE; PS00078; COX2; 1. Copper; Electron transport; Inner membrane; Membrane; Oxidoreductase; Respiratory chain; Transmembrane; Transport; Mitochondrion. SEQUENCE 253 AA; 29095 MW; BBE9F70ED865C52 CRC64;  
 Query Match 87.8%; Score 36; DB 8; Length 253;  
 Best Local Similarity 85.7%; Pred. No. 16;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 HYAKNPI 7  
 DB 64 HYKNPI 70

GN COX2.  
 OS Amphicarpaea bracteata (Hog peanut).  
 OG Mitochondrion.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;  
 OC Amphicarpaea.  
 OC NCBI\_TaxID=45679;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20040642; PubMed=10570164;  
 RA Adams K.L., Song K., Roessler P.G., Nugent J.M., Doyle J.L.,  
 RA Doyle J.J., Palmer J.D.;  
 RT "Intracellular gene transfer in action: dual transcription and multiple silencings of nuclear and mitochondrial cox2 genes in legumes.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 96:13863-13868 (1999).  
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2 TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER A CENTER TO THE BIMETALLIC CENTER OF THE CATALYTIC SUBUNIT 1 (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 4 FERRICYTOCHROME C + 2 H(2)O.  
 CC -1- COFACTOR: COPPER A (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL INNER MEMBRANE (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY. EMBL; AF207682; AAF43639.1; InterPro; IPR001505; Copper\_CuA. InterPro; IPR002429; Cyt\_c\_ox\_2. Pfam; PF00116; COX2; 1. PRINTS; PR01166; CYCOXIDASEII. PRODOM; PD000131; Copper\_CuA; 1. PROSITE; PS00078; COX2; 1. Copper; Electron transport; Inner membrane; Membrane; Oxidoreductase; Respiratory chain; Transmembrane; Transport; Mitochondrion. SEQUENCE 253 AA; 29033 MW; B6A4380C75219B90 CRC64;  
 Query Match 87.8%; Score 36; DB 8; Length 253;  
 Best Local Similarity 85.7%; Pred. No. 16;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 HYAKNPI 7  
 DB 64 HYKNPI 70  
 RESULT 14  
 Q9MM59 PRELIMINARY; PRT; 253 AA.  
 AC Q9MM59;  
 DT 01-OCT-2000 (TRENBLrel. 15, Created)  
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
 DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)  
 DE Cytochrome oxidase subunit 2 (EC 1.9.3.1) (Cytochrome c oxidase polypeptide II).  
 GN Pseudominia comosa.  
 OS Mitochondrion.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;  
 OC Pseudominia.  
 OC NCBI\_TaxID=45681;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20040642; PubMed=10570164;  
 RA Adams K.L., Song K., Roessler P.G., Nugent J.M., Doyle J.L.,  
 RA Doyle J.J., Palmer J.D.;  
 RT "Intracellular gene transfer in action: dual transcription and multiple silencings of nuclear and mitochondrial cox2 genes in legumes.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 96:13863-13868 (1999).  
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2 TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER A CENTER TO THE BIMETALLIC CENTER OF THE CATALYTIC SUBUNIT 1 (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 4 FERRICYTOCHROME C + 2 H(2)O.  
 CC -1- COFACTOR: COPPER A (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL INNER MEMBRANE (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY. EMBL; AF207682; AAF43639.1; InterPro; IPR001505; Copper\_CuA. InterPro; IPR002429; Cyt\_c\_ox\_2. Pfam; PF00116; COX2; 1. PRINTS; PR01166; CYCOXIDASEII. PRODOM; PD000131; Copper\_CuA; 1. PROSITE; PS00078; COX2; 1. Copper; Electron transport; Inner membrane; Membrane; Oxidoreductase; Respiratory chain; Transmembrane; Transport; Mitochondrion. SEQUENCE 253 AA; 29033 MW; B6A4380C75219B90 CRC64;  
 Query Match 87.8%; Score 36; DB 8; Length 253;  
 Best Local Similarity 85.7%; Pred. No. 16;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 HYAKNPI 7  
 DB 64 HYKNPI 70



RT multiple silencings of nuclear and mitochondrial cox2 genes in  
 RT legumes.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 96:13863-13868(1999).  
 CC -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY  
 CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-  
 CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2  
 CC TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER  
 CC A CENTER TO THE BINUCLEAR CENTER OF THE CATALYTIC SUBUNIT 1 (BY  
 CC SIMILARITY).  
 CC -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 4 FERRICYTOCHROME  
 CC C + 2 H(2)O.  
 CC -!- COFACTOR: COPPER A (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL  
 CC INNER MEMBRANE (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.  
 DR EMBL; AF207674; AAF43632.1; -.  
 DR InterPro; IPR001505; Copper\_CuA.  
 DR InterPro; IPR002429; Cyt\_c\_ox\_2.  
 DR Pfam; PF00116; COX2; 1. \_ox\_2.  
 DR Pfam; PF02790; COX2; 1. \_ox\_2.  
 DR PRINTS; PR01166; CYCOXIDASEII.  
 DR ProDom; PD000131; Copper\_CuA; 1.  
 KW Copper; Electron transport; Inner membrane; Membrane; Oxidoreductase;  
 KW Respiratory chain; Transmembrane; Transport; Mitochondrion.  
 SQ SEQUENCE 253 AA; 28820 MW; 018E680BC17D9AC0 CRC64;

Query Match 87.8%; Score 36; DB 8; Length 253;  
 Best Local Similarity 85.7%; Pred. No. 16;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 HYAKNPI 7  
 |||||  
 DB 64 HYKKNPI 70

## RESULT 15

Q9NM55 PRELIMINARY; PRT; 253 AA.  
 AC Q9NM55;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE Cytochrome oxidase subunit 2 (EC 1.9.3.1) (Cytochrome c oxidase  
 DE polypeptide II).  
 GN COX2.  
 OS Pueraria phaseoloides.  
 OG Mitochondrion.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Pueraria.  
 OX NCBI\_TaxID=109224;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=2040642; PubMed=10570164;  
 RA Adams K.L., Song K., Roessler P.G., Nugent J.M., Doyle J.L.,  
 RA Doyle J.J., Palmer J.D.;  
 RT "Intracellular gene transfer in action: dual transcription and  
 RT multiple silencings of nuclear and mitochondrial cox2 genes in  
 RT legumes.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 96:13863-13868(1999).  
 CC -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY  
 CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNIT 1-  
 CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. SUBUNIT 2  
 CC TRANSFERS THE ELECTRONS FROM CYTOCHROME C VIA ITS BINUCLEAR COPPER  
 CC A CENTER TO THE BINUCLEAR CENTER OF THE CATALYTIC SUBUNIT 1 (BY  
 CC SIMILARITY).  
 CC -!- CATALYTIC ACTIVITY: 4 FERROCYTOCHROME C + O(2) = 4 FERRICYTOCHROME  
 CC C + 2 H(2)O.  
 CC -!- COFACTOR: COPPER A (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL  
 CC INNER MEMBRANE (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME C OXIDASE SUBUNIT 2 FAMILY.  
 DR EMBL; AF207678; AAF43636.1; -.

DR InterPro; IPR001505; Copper\_CuA.  
 DR InterPro; IPR002429; Cyt\_c\_ox\_2.  
 DR Pfam; PF00116; COX2; 1.  
 DR PRINTS; PR01166; CYCOXIDASEII.  
 DR ProDom; PD000131; Copper\_CuA; 1.  
 KW Copper; Electron transport; Inner membrane; Membrane; Oxidoreductase;  
 KW Respiratory chain; Transmembrane; Transport; Mitochondrion.  
 SQ SEQUENCE 253 AA; 28812 MW; 4B8F9E41C1097000 CRC64;  
 Query Match 87.8%; Score 36; DB 8; Length 253;  
 Best Local Similarity 85.7%; Pred. No. 16;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 OY 1 HYAKNPI 7  
 |||||  
 DB 64 HYKKNPI 70  
 Search completed: February 11, 2004, 17:09:18  
 Job time : 27.5833 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 16:51:43 ; Search time 5.16667 Seconds  
(without alignments)  
63.714 Million cell updates/sec

Title: US-09-901-187C-2

Perfect score: 41

Sequence: 1 HYAKNPI 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	36	87.8	258	1	COX2_PEA
2	36	87.8	260	1	COX2_BETVU
3	36	87.8	260	1	COX2_SOYEN
4	34	82.9	551	1	Y275_HAEIN
5	33	80.5	251	1	COX2_WARPO
6	33	80.5	585	1	PTNB_MOUSE
7	33	80.5	593	1	PTNB_HUMAN
8	33	80.5	593	1	PTNB_RAT
9	33	80.5	2109	1	RRPL_VSVJO
10	33	80.5	2109	1	RRPL_VSVSJ
11	31	75.6	213	1	GPH_AQUAE
12	31	75.6	352	1	OPSF_ANGAN
13	31	75.6	372	1	LEM1_MACMU
14	31	75.6	372	1	LEM1_PAPHA
15	31	75.6	453	1	TBD_HUMAN
16	31	75.6	455	1	TBD_MOUSE
17	31	75.6	600	1	MCW3_MAIZE
18	31	75.6	814	1	POL_IPNAI
19	31	75.6	833	1	CM41_YEAST
20	31	75.6	867	1	POL_IPWA
21	31	75.6	943	1	YLVE_CABEL
22	31	75.6	1054	1	RPOC_WEIPA
23	30	73.2	35	1	Y847_BORBU
24	30	73.2	154	1	YKYB_BACSU
25	30	73.2	247	1	COX2_KLUTH
26	30	73.2	251	1	COX2_KLUTH
27	30	73.2	251	1	COX2_YEAST
28	30	73.2	258	1	COX2_OENBE
29	30	73.2	312	1	Y351_BUCAP
30	30	73.2	387	1	AMP1_YEAST
31	30	73.2	388	1	GUN3_HUMAN
32	30	73.2	394	1	AMP1_HUMAN
33	30	73.2	412	1	AMAT_CALCIN

#### ALIGNMENTS

##### RESULT 1

##### COX2\_PEA

ID\_COX2\_PEA STANDARD; PRT; 258 AA.

AC P08744; 30 73.2 414 1 DP41\_BACSU  
DT 01-AUG-1988 (Rel. 08, Created) 477 1 P3\_HUMAN  
DT 01-NOV-1995 (Rel. 32, Last sequence update) 524 1 C6G1\_DROME  
DT 28-FEB-2003 (Rel. 41, Last annotation update) 541 1 MASY\_MYXXA  
DE Cytochrome c oxidase polypeptide II (EC 1.9.3.1). 605 1 APW2\_YEAST  
GN COX2 OR COII. 612 1 EXO2\_BPTS  
OS Pisum sativum (Garden pea). 845 1 CSW\_DROME  
OG Mitochondrion. 875 1 NPP3\_HUMAN  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; 875 1 NPP3\_RAT  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; 1057 1 TLD\_DROME  
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum. 1237 1 CT26\_HUMAN  
OX NCBI\_TaxID:3888; 1237 1 RRPL\_TSWV1  
RN [1] 2875 1  
RP SEQUENCE FROM N.A. 2875 1  
RX MEDLINE-85215679; PubMed-2987876; 2875 1  
RA Moon E., Kao T.-H., Wu R.; 2875 1  
RT "Pea cytochrome oxidase subunit II gene has no intron and generates 2875 1  
RL two mRNA transcripts with different 5'-termini."; 2875 1  
CC Nucleic Acids Res. 13:3185-3212(1985). 2875 1  
CC -!- FUNCTION: Cytochrome c oxidase is the component of the respiratory 2875 1  
CC chain that catalyzes the reduction of oxygen to water. Subunits 1- 2875 1  
CC 3 form the functional core of the enzyme complex. Subunit 2 2875 1  
CC transfers the electrons from cytochrome c via its binuclear copper 2875 1  
CC A center to the bimetallic center of the catalytic subunit 1. 2875 1  
CC -!- CATALYTIC ACTIVITY: 4 ferrocytochrome c + O(2) = 4 ferricytochrome 2875 1  
CC c + 2 H(2)O. 2875 1  
CC -!- COFACTOR: Copper A. 2875 1  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial 2875 1  
CC inner membrane. 2875 1  
CC -!- SIMILARITY: Belongs to the cytochrome c oxidase subunit 2 family. 2875 1  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration 2875 1  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation - 2875 1  
CC the European Bioinformatics Institute. There are no restrictions on its 2875 1  
CC use by non-profit institutions as long as its content is in no way 2875 1  
CC modified and this statement is not removed. Usage by and for commercial 2875 1  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> 2875 1  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)). 2875 1  
CC  
CC EMBL; X02433; CAA26282.1; - 2875 1  
CC PIR; A23012; OBPM2. 2875 1  
CC InterPro; IPR001505; Copper\_CuA. 2875 1  
CC InterPro; IPR002429; Cyt\_c\_ox\_2. 2875 1  
CC Pfam; PF00116; COX2\_1. 2875 1  
CC Pfam; PF02790; COX2\_TM; 1 2875 1  
CC PRINTS; PR01166; CYCOXIDASEII. 2875 1  
CC PRODOM; PD000131; Copper\_CuA; 1. 2875 1  
CC PROSITE; PS00078; COX2; FALSE NEG. 2875 1  
CC Oxidoreductase; Copper; Mitochondrion; Transmembrane; Inner membrane; 2875 1  
KW Electron transport; Respiratory chain. 2875 1  
FT DOMAIN 1 39 MITOCHONDRIAL INTERMEMBRANE (POTENTIAL). 2875 1  
FT TRANSMEM 40 61 POTENTIAL. 2875 1  
FT DOMAIN 62 85 MITOCHONDRIAL MATRIX (POTENTIAL). 2875 1

Search completed: February 11, 2004, 17:54:06  
Job time : 24.6667 secs

ORGANISM: Mycobacterium tuberculosis  
US-10-080-170-604  
Query Match 85.3%; Score 29; DB 16; Length 264;  
Best Local Similarity 71.4%; Pred. No. 2.9e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 1 RRRGMAI 7  
Db 17 RRRGLAL 23  
RESULT 15  
US-10-084-843-71  
Sequence 71, Application US/10084843  
Publication No. US20030143243A1  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
Skeiky, Yasir A.W.  
Dillon, Davin C.  
Campos-Neto, Antonio  
Houghton, Raymond  
Vedvick, Thomas S.  
Twardzik, Daniel R.  
Lodes, Michael J.  
Hendrickson, Ronald C.  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY  
AND DIAGNOSIS OF TUBERCULOSIS  
NUMBER OF SEQUENCES: 355  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED and BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/084,843  
FILING DATE: 25-Feb-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/072,967  
FILING DATE: 05-MAY-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.411C9  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 71:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 267 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 71:  
US-10-084-843-71  
Query Match 85.3%; Score 29; DB 12; Length 267;  
Best Local Similarity 71.4%; Pred. No. 3e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 1 RRRGMAI 7  
Db 20 RRRGLAL 26

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 16:56:09 ; Search time 10.4167 Seconds  
(without alignments)  
28.433 Million cell updates/sec

Title: US-09-901-187C-4  
Perfect score: 34  
Sequence: 1 RRRGNM 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A-COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B-COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A-COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B-COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS-COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	31	91.2	242	4	US-09-252-991A-19021
2	30	88.2	191	4	US-09-252-991A-19742
3	30	88.2	205	4	US-09-252-991A-29378
4	30	88.2	353	4	US-09-252-991A-22306
5	30	88.2	1002	4	US-09-252-991A-27980
6	30	88.2	1299	4	US-09-252-991A-31121
7	29	85.3	267	3	US-08-818-112-71
8	29	85.3	267	4	US-08-818-111-72
9	29	85.3	267	4	US-09-056-556-71
10	29	85.3	267	4	US-09-072-596-72
11	29	85.3	284	4	US-09-056-556-162
12	29	85.3	284	4	US-09-072-596-157
13	29	85.3	474	4	US-09-252-991A-22590
14	29	85.3	482	4	US-09-252-991A-30370
15	29	85.3	485	4	US-09-252-991A-25242
16	29	85.3	584	4	US-09-252-991A-24311
17	29	85.3	1021	4	US-09-252-991A-27405
18	28	82.4	177	4	US-09-252-991A-25407
19	28	82.4	278	3	US-08-522-813-4
20	28	82.4	293	1	US-08-446-925-5
21	28	82.4	293	2	US-09-146-331-5
22	28	82.4	293	2	US-08-896-885-5
23	28	82.4	293	4	US-09-375-256-5
24	28	82.4	293	4	US-09-561-756-21
25	28	82.4	293	4	US-09-227-721-21
26	28	82.4	293	4	US-08-983-502-31
27	28	82.4	293	4	US-09-376-156-5

28	28	82.4	293	4	US-08-724-378D-6	Sequence 6, Appl
29	28	82.4	293	4	US-09-516-747-31	Sequence 31, Appl
30	28	82.4	293	5	PCT-US96-10521-31	Sequence 31, Appl
31	28	82.4	297	4	US-09-252-991A-18170	Sequence 18170, A
32	28	82.4	300	4	US-09-561-756-36	Sequence 36, Appl
33	28	82.4	300	4	US-09-227-721-36	Sequence 36, Appl
34	28	82.4	302	4	US-08-311-731A-206	Sequence 206, App
35	28	82.4	341	4	US-09-252-991A-27327	Sequence 27327, A
36	28	82.4	363	3	US-09-301-665-4	Sequence 4, Appl
37	28	82.4	460	4	US-09-252-991A-24811	Sequence 24811, A
38	28	82.4	712	4	US-09-252-991A-20471	Sequence 20471, A
39	27	79.4	107	4	US-09-252-991A-30944	Sequence 30944, A
40	27	79.4	145	4	US-09-252-991A-23915	Sequence 23915, A
41	27	79.4	161	4	US-09-252-991A-17481	Sequence 17481, A
42	27	79.4	162	4	US-09-252-991A-24838	Sequence 24838, A
43	27	79.4	206	4	US-09-252-991A-16775	Sequence 16775, A
44	27	79.4	209	4	US-09-252-991A-28575	Sequence 28575, A
45	27	79.4	221	4	US-09-252-991A-20588	Sequence 20588, A

## ALIGNMENTS

RESULT 1  
US-09-252-991A-19021  
; Sequence 19021, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252.991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 19021  
; LENGTH: 242  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-19021

Query Match 91.2%; Score 31; DB 4; Length 242;  
Best Local Similarity 85.7%; Pred. No. 20;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRRGNM 7  
|||||:  
Db 145 RRRGNM 151

RESULT 2  
US-09-252-991A-19742  
; Sequence 19742, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252.991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 19742  
; LENGTH: 191  
; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-19742

Query Match 88.2%; Score 30; DB 4; Length 191;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRRGWA 6  
|:|||||  
DB 16 RRRGWA 21

## RESULT 3

US-09-252-991A-29378  
; Sequence 29378, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 29378  
; LENGTH: 205  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-29378

Query Match 88.2%; Score 30; DB 4; Length 205;  
Best Local Similarity 100.0%; Pred. No. 28;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRRGWA 6  
|:|||||  
DB 38 RRRGWA 43

## RESULT 4

US-09-252-991A-22306  
; Sequence 22306, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 22306  
; LENGTH: 353  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-22306

Query Match 88.2%; Score 30; DB 4; Length 353;  
Best Local Similarity 85.7%; Pred. No. 49;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRRGMAI 7  
|:|||||  
DB 37 RRRGMAI 43

## RESULT 5

US-09-252-991A-27980  
; Sequence 27980, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 27980  
; LENGTH: 1002  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-27980

Query Match 88.2%; Score 30; DB 4; Length 1002;  
Best Local Similarity 71.4%; Pred. No. 1.4e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRRGMAI 7  
|:|||||  
DB 421 RRRGLAV 427

## RESULT 6

US-09-252-991A-31121  
; Sequence 31121, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 31121  
; LENGTH: 1299  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-31121

Query Match 88.2%; Score 30; DB 4; Length 1299;  
Best Local Similarity 71.4%; Pred. No. 1.9e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRRGMAI 7  
|:|||||  
DB 928 RRRGLAV 934

## RESULT 7

US-08-818-112-71  
; Sequence 71, Application US/08818112  
; Patent No. 6290969  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillion, Davin C.  
; APPLICANT: Campos-Neto, Antonio  
; APPLICANT: Houghton, Raymond

```

; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 71:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 267 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-818-112-71

```

```

Query Match      85.3%; Score 29; DB 3; Length 267;
Best Local Similarity 71.4%; Pred. No. 61;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 RRRGMJ 7
        |||||:
        20 RRRGLAL 26

```

```

; RESULT 8
; US-08-818-111-72
; Sequence 72, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,111

```

```

; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 72:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 267 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-818-111-72

```

```

Query Match      85.3%; Score 29; DB 4; Length 267;
Best Local Similarity 71.4%; Pred. No. 61;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 RRRGMJ 7
        |||||:
        20 RRRGLAL 26

```

```

; RESULT 9
; US-09-056-556-71
; Sequence 71, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 71:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 267 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-056-556-71

```

```

Query Match      85.3%; Score 29; DB 4; Length 267;
Best Local Similarity 71.4%; Pred. No. 61;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 RRRGMJ 7
        |||||:

```

TREAT

Db 20 RRRGLAL 26

RESULT 10  
US-09-072-596-72  
; Sequence 72, Application US/09072596  
; Patent No. 6458366  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Campos-Neto, Antonia  
; APPLICANT: Houghton, Raymond  
; APPLICANT: Vedvick, Thomas S.  
; APPLICANT: Twardzik, Daniel R.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Hendrickson, Ronald C.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF  
; NUMBER OF SEQUENCES: 350  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/072,596  
; FILING DATE: 05-MAY-1998  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.417C9  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-6031  
; TELEFAX: (206) 622-4900  
; INFORMATION FOR SEQ ID NO: 72:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 267 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-09-072-596-72

Query Match 85.3%; Score 29; DB 4; Length 267;  
Best Local Similarity 71.4%; Pred. No. 61;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRRGMAI 7  
Db 20 RRRGLAL 26

RESULT 11  
US-09-056-556-162  
; Sequence 162, Application US/09056556  
; Patent No. 6350456  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, Davin C.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND  
; NUMBER OF SEQUENCES: 241  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle

TREATM

STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/072,596  
FILING DATE: 05-MAY-1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.457  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 622-6031  
INFORMATION FOR SEQ ID NO: 162:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 284 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
US-09-056-556-162

Query Match 85.3%; Score 29; DB 4; Length 284;  
Best Local Similarity 71.4%; Pred. No. 65;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRRGMAI 7  
Db 257 RRRGVAV 263

RESULT 12  
US-09-072-596-157  
; Sequence 157, Application US/09072596  
; Patent No. 6458366  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Campos-Neto, Antonia  
; APPLICANT: Houghton, Raymond  
; APPLICANT: Vedvick, Thomas S.  
; APPLICANT: Twardzik, Daniel R.  
; APPLICANT: Hendrickson, Ronald C.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF  
; NUMBER OF SEQUENCES: 350  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98104-7092  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/072,596  
; FILING DATE: 05-MAY-1998  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Maki, David J.  
; REGISTRATION NUMBER: 31,392  
; REFERENCE/DOCKET NUMBER: 210121.417C9  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-6031  
; TELEFAX: (206) 622-4900  
; INFORMATION FOR SEQ ID NO: 72:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 267 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-09-072-596-72

Query Match 85.3%; Score 29; DB 4; Length 267;  
Best Local Similarity 71.4%; Pred. No. 61;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRRGMAI 7  
Db 20 RRRGLAL 26

RESULT 11  
US-09-056-556-162  
; Sequence 162, Application US/09056556  
; Patent No. 6350456  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Dillon, Davin C.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND  
; NUMBER OF SEQUENCES: 241  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SEED and BERRY LLP  
; STREET: 6300 Columbia Center, 701 Fifth Avenue  
; CITY: Seattle

TREATM

TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031  
INFORMATION FOR SEQ ID NO: 157:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 284 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
US-09-072-596-157

Query Match 85.3%; Score 29; DB 4; Length 284;  
Best Local Similarity 71.4%; Pred. No. 65;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRRGMAI 7  
| | | | |  
Db 257 RRRGVAV 263

## RESULT 13

US-09-252-991A-22590  
; Sequence 22590, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 22590  
; LENGTH: 474  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-22590

Query Match 85.3%; Score 29; DB 4; Length 474;  
Best Local Similarity 71.4%; Pred. No. 1.1e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRRGMAI 7  
| | | | |  
Db 215 QRRGMV 221

## RESULT 14

US-09-252-991A-30970  
; Sequence 30970, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 30970  
; LENGTH: 482  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-30970

Query Match 85.3%; Score 29; DB 4; Length 482;

Best Local Similarity 71.4%; Pred. No. 1.1e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 1 RRRGMAI 7  
| | | | |  
Db 449 RRRGIIV 455

## RESULT 15

US-09-252-991A-25242  
; Sequence 25242, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 25242  
; LENGTH: 485  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-25242

Query Match 85.3%; Score 29; DB 4; Length 485;  
Best Local Similarity 71.4%; Pred. No. 1.1e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRRGMAI 7  
| | | | |  
Db 233 RRRGIIV 239

Search completed: February 11, 2004, 17:13:34  
Job time : 10.4167 secs



GenCore version 5.1.6  
Copyright (C) 1993 - 2004 Compugen Ltd.

QM protein - protein search, using sw model

Run on: February 11, 2004, 16:51:33 ; Search time 32.25 Seconds  
(without alignments)  
34.452 Million cell updates/sec

Title: US-09-901-187C-4  
Perfect score: 34  
Sequence: 1 RRRGMAI 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A.Geneseq.19Jun03.\*  
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	100.0	7	AAE14549	Human alpha-synuclein
2	30	88.2	112	ABP32807	Human nucleasase-lik
3	30	88.2	130	ABP63963	Human ORF333. Hom
4	30	88.2	162	AAU49274	Propionibacterium
5	30	88.2	263	ABU05562	M. tuberculosis an
6	30	88.2	477	AAU51112	Propionibacterium
7	30	88.2	715	ABJ25717	Aspergillus fumiga
8	30	88.2	715	ABJ26317	Aspergillus fumiga
9	30	88.2	1335	ABP71593	Drosophila melanog

10	29	85.3	95	24	ABU00291	Human novel polype
11	29	85.3	129	20	AAV48598	Human breast tumou
12	29	85.3	264	23	ABU05953	M. tuberculosis an
13	29	85.3	267	18	AAW32427	Mycobacterium tube
14	29	85.3	267	18	AAW32359	Mycobacterium tube
15	29	85.3	267	19	AAW81862	M. tuberculosis im
16	29	85.3	267	19	AAW64299	Mycobacterium tube
17	29	85.3	267	20	AAV39101	M. tuberculosis an
18	29	85.3	267	20	AAV38964	M. tuberculosis re
19	29	85.3	284	19	AAW81733	Mycobacterium tube
20	29	85.3	284	19	AAW64366	M. tuberculosis an
21	29	85.3	284	20	AAV39163	M. tuberculosis re
22	29	85.3	284	20	AAV39020	Human polypeptide
23	29	85.3	611	22	AAW41662	Novel human diagn
24	29	85.3	703	22	ABG17698	Candida albicans e
25	29	85.3	725	23	ABP73689	Human polypeptide
26	29	85.3	922	23	AAU76150	Rice lipoxigenase
27	28	82.4	49	22	AAO06965	Human polypeptide
28	28	82.4	64	22	AAU56222	Propionibacterium
29	28	82.4	70	22	AAU56770	Propionibacterium
30	28	82.4	70	22	AAU54122	Human polypeptide
31	28	82.4	135	22	AAO04393	Propionibacterium
32	28	82.4	175	22	AAU60435	Human ovarian anti
33	28	82.4	231	23	ABP42451	Arabidopsis thalia
34	28	82.4	242	21	AAU23229	Microspora car
35	28	82.4	252	22	ABP06906	Orthomycospora car
36	28	82.4	252	24	ABP99287	Streptomyces virid
37	28	82.4	258	24	ABP76723	Interleukin 1 conv
38	28	82.4	278	21	AAV85061	Apoptotic cysteine
39	28	82.4	293	18	AAW06244	Amino acid sequenc
40	28	82.4	293	20	AAV21720	Human caspase-6.
41	28	82.4	293	22	AAE00603	Human caspase 6 re
42	28	82.4	293	23	AAO21925	Human caspase-6 SE
43	28	82.4	293	23	ABJ01221	Signal peptide-pro
44	28	82.4	299	10	AAV94148	Amino acid sequenc
45	28	82.4	300	20	AAV21727	

ALIGNMENTS

RESULT 1  
AAE14549  
ID AAE14549 standard; peptide; 7 AA.

AC AAE14549;

XX DT 17-MAY-2002 (first entry)

XX DE Human alpha-synuclein aggregation inhibitor #4.

XX KW Alpha-synuclein; inhibitor; neurodegenerative disease; Lewy body;  
XX KW Parkinson's disease; Alzheimer's disease; diffuse Lewy body disease;  
XX KW multiple system atrophy; Hallervorden-Spatz disease; human.

XX OS Homo-sapiens

XX PN WO200204482-A1.

XX PD 17-JAN-2002.

XX PF 06-JUN-2001; 2001WO-US21379.

XX PR 07-JUL-2000; 2000US-217319P.

XX PR 28-MAR-2001; 2001US-279199P.

XX PA (PANA-) PANACEA PHARM INC.

XX PI Wlozoin B, Osretova-Golts N, Lebowitz MS;

XX DR WPI; 2002-179695/23.

XX PT Determination of an agent capable of inhibiting aggregation of alpha

PT synuclein useful for treating a neurodegenerative disease involves  
PT determining aggregation of alpha synuclein in the presence of exogenous  
PT iron or copper  
XX  
XX Claim 40; Page 37; 52pp; English.  
XX  
CC The invention relates to screening of inhibitors of alpha-synuclein  
CC aggregation in the presence of exogenous iron or copper. The inhibitors  
CC are magnesium and alpha-synuclein binding peptides, which are  
CC useful for treating neurodegenerative disease that involves  
CC the formation of Lewy bodies e.g. Parkinson's disease (PD), Alzheimer's  
CC disease (AD), diffuse Lewy body disease, mixed AD-PD, multiple system  
CC atrophy and Hallervorden-Spatz disease. The present sequence is a  
CC peptide that binds to C-terminal portion of human alpha-synuclein and  
CC inhibits its aggregation.  
XX  
XX Sequence 7 AA;  
XX  
XX Query Match 100.0%; Score 34; DB 23; Length 7;  
XX Best Local Similarity 100.0%; Pred. No. 9.3e+05;  
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
XX QY 1 RRRGMAI 7  
XX | | | | |  
XX Db 1 RRRGMAI 7  
XX  
XX RESULT 2  
XX ID ABP32807 standard; Protein; 112 AA.  
XX AC ABP32807;  
XX  
XX 08-JUL-2002 (first entry)  
XX  
XX Human nuclease-like ORF1780 protein, SEQ ID NO:3560.  
XX  
XX Human; ORF; open reading frame; ORFX; drug screening; diagnosis;  
XX disease monitoring; cytokine; cell proliferation; cell differentiation;  
XX immune modulation; haematopoiesis regulation; tissue growth;  
XX angiogenesis; activin; inhibin; chemotactic; chemokinetic; haemostatic;  
XX thrombolytic; tumour inhibition; bodily characteristic; fertility;  
XX behaviour; cancer; proliferative disorder; neurological disorder;  
XX cardiovascular disease; immune system disorder; organ transplantation;  
XX tissue growth disorder; tissue regeneration disorder; diabetes mellitus;  
XX hypothyroidism; cholesterol ester storage disease; infection; vulnary;  
XX vasotropic; antipsoriatic; antidiabetic; cytostatic; neutropic;  
XX neuroprotective; antiatherosclerotic; anticoagulant; immunomodulator;  
XX cardiant; hypotensive; antithyroid; antiinflammatory; immunomodulator;  
XX dermatological; analgesic; virucide; antibacterial; fungicide.  
XX  
XX Homo sapiens.  
XX  
XX WO200190366-A2.  
XX  
XX 29-NOV-2001.  
XX  
XX 24-MAY-2001; 2001WO-US17075.  
XX  
XX 24-MAY-2000; 2000US-206690P.  
XX  
XX (CURA-) CURAGEN CORP.  
XX  
XX Leach MD, Shimkets RA;  
XX  
XX WPI; 2002-106200/14.  
XX N-PSDB; AEN76833.  
XX  
XX Novel human polypeptides and polynucleotides useful for diagnosing,  
XX preventing and treating cardiovascular disease, neurodegenerative,  
XX hyperproliferative disorders and disorders related to organ  
XX transplantation  
XX

PS Claim 10; Page 1150; 2508pp; English.  
XX  
XX Sequences ABP31028-ABP35561 represent 4534 novel human proteins  
XX designated ORF (Open reading frame) 1-4534, and sequences ABN75054-  
XX ABN79587 represent cDNAs encoding them. The invention also encompasses  
XX polypeptides at least 80% identical to the ORF1-ORF4534 (collectively  
XX referred to as ORFX) proteins, polynucleotides at least 85% identical to  
XX the ORFX nucleic acid sequences, vectors and host cells comprising ORFX  
XX polynucleotides, the recombinant production of ORFX proteins, antibodies  
XX specific for ORFX proteins, methods of detecting ORFX polynucleotides and  
XX polypeptides, methods of screening for modulators of ORFX expression or  
XX activity, and methods of screening individuals for a predisposition to an  
XX ORFX-associated disorder. The ORFX proteins of the invention have a wide  
XX range of biological activities, such as cytokine, cell proliferation,  
XX cell differentiation, immune modulation, haematopoiesis regulation,  
XX tissue growth, angiogenesis, activin or inhibin activity, chemotactic/  
XX chemokinetic activity, haemostatic activity, thrombolytic activity,  
XX receptor/ligand, antiinflammatory activity, tumour inhibition activity,  
XX of bodily characteristics, fertility and behaviour. ORFX proteins,  
XX nucleic acids and antibodies may be used in the treatment of cancers,  
XX other proliferative disorders such as psoriasis and benign tumours,  
XX neurological disorders such as epilepsy and Alzheimer's disease,  
XX cardiovascular diseases, immune system disorders, disorders related to  
XX organ transplantation, disorders of tissue growth and regeneration,  
XX diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester  
XX storage disease, and infectious diseases caused by viral, bacterial,  
XX fungal and other pathogens. ORFX nucleic acids may also be used as a  
XX source of primers and probes, in the detection of ORFX genomic sequences  
XX or transcripts, in the identification and cloning of homologous  
XX sequences, in genetic diagnosis, and in forensic biology. The ORFX  
XX nucleic acids may additionally be used to produce transgenic animals  
XX which may be useful for studying the function and/or activity of ORFX  
XX protein, and in drug screening. The ORFX proteins may also be used as  
XX immunogens to generate specific antibodies, which are useful in the  
XX diagnosis, treatment and monitoring of ORFX-associated diseases.  
XX  
XX Sequence 112 AA;  
XX  
XX Query Match 88.2%; Score 30; DB 23; Length 112;  
XX Best Local Similarity 71.4%; Pred. No. 71;  
XX Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
XX  
XX QY 1 RRRGMAI 7  
XX | | | | |  
XX Db 71 KRRGMV 77  
XX  
XX RESULT 3  
XX ABP3963  
XX ID ABP3963 standard; Protein; 130 AA.  
XX AC ABP3963;  
XX  
XX 04-NOV-2002 (first entry)  
XX  
XX Human ORF333.  
XX  
XX Cytostatic; Cardiant; Anti-allergic; Immunosuppressive; Vulnary;  
XX Antiinflammatory; gene therapy; human; ORFX; atherogenic; platelet;  
XX human umbilical vein endothelial cell; HUVEC; atherosclerotic plaque;  
XX cancer; cardiovascular disease; allergy; autoimmune disease;  
XX wound healing; blood coagulation disorder; inflammatory disorder.  
XX  
XX Homo sapiens.  
XX  
XX US2002082206-A1.  
XX  
XX 27-JUN-2002.  
XX  
XX 30-MAY-2001; 2001US-0867550.  
XX  
XX 30-MAY-2000; 2000US-208427P.  
XX



XX Identifying and selecting genes for survival or virulence of  
PT Mycobacterium tuberculosis and M. leprae -  
XX  
XX Claim 17; Page 392-393; 874pp; English.  
XX  
XX This invention relates to a novel method for identifying essential genes  
CC for survival or virulence of mycobacteria species. The method comprises  
CC aligning the genomic sequence of a first mycobacterium species on a  
CC genomic sequence of a second mycobacterium species and selecting a  
CC polynucleotide sequence that is highly conserved in both genomes with no  
CC counterparts in other bacterial genomic sequences and that corresponds  
CC to an essential gene for the survival or virulence of mycobacterium  
CC species. The method of the invention is useful for detecting M.  
CC tuberculosis or M. leprae infection. The method reduces the number of  
CC potential new targets and protective antigens for new drugs and vaccine  
CC compositions to treat and prevent mycobacterial diseases, particularly  
CC tuberculosis and leprosy. The present sequence represents a marker  
CC protein from Mycobacterium tuberculosis and Mycobacterium leprae  
CC identified using the method of the invention.  
XX  
XX SQ Sequence 263 AA;

Query Match 88.2%; Score 30; DB 23; Length 263;  
Best Local Similarity 71.4%; Pred. No. 1.6e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 1 RRRGMAI 7  
Db 17 RRRGLAV 23

RESULT 6  
AAUS1112  
ID AAUS1112 standard; Protein; 477 AA.  
XX  
XX AC AAUS1112;  
XX  
XX DT 27-FEB-2002 (first entry)  
XX  
XX DE Propionibacterium acnes immunogenic protein #12008.  
XX  
XX KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
XX uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
XX inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
XX dermatological; osteopathic; neuroprotectant.  
XX  
XX OS Propionibacterium acnes.  
XX  
XX PN WO200191581-A2.  
XX  
XX PD 01-NOV-2001.  
XX  
XX PF 20-APR-2001; 2001WO-US12865.  
XX  
XX PR 21-APR-2000; 2000US-199047P.  
XX  
XX PR 02-JUN-2000; 2000US-208841P.  
XX  
XX PR 07-JUL-2000; 2000US-216747P.  
XX  
XX PA (CORI-) CORIXA CORP.  
XX  
XX PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
XX L' Maisonneuve J, Zhang Y, Jen S, Carter D;  
XX  
XX DR WPI; 2001-616774/71.  
XX  
XX DR N-PSDB; AAS59549.  
XX  
XX PT Propionibacterium acnes polypeptides and nucleic acids useful for  
XX vaccinating against and diagnosing infections, especially useful for  
XX treating acne vulgaris -  
XX  
XX Example 1; SEQ ID No 12307; 1069pp; English.

XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
CC polypeptides. The proteins and their associated DNA sequences are used in  
CC the treatment, prevention and diagnosis of medical conditions caused by  
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
CC P. acnes is also involved in infections of bone, joints and the central  
CC nervous system, however it is particularly involved in the inflammatory  
CC lesions associated with acne vulgaris. A method for detecting the  
CC presence or absence of P. acnes in a patient comprises contacting a  
CC sample with a binding agent that binds to the proteins of the invention  
CC and determining the amount of bound protein in the sample. The  
CC polypeptides may be used as antigens in the production of antibodies  
CC specific for P. acnes proteins. These antibodies can be used to  
CC downregulate expression and activity of P. acnes polypeptides and  
CC therefore treat P. acnes infections. The antibodies may also be used as  
CC diagnostic agents for determining P. acnes presence, for example, by  
CC enzyme linked immunosorbent assay (ELISA).  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX SQ Sequence 477 AA;

Query Match 88.2%; Score 30; DB 22; Length 477;  
Best Local Similarity 71.4%; Pred. No. 2.8e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 1 RRRGMAI 7  
Db 214 KRRGMVAV 220

RESULT 7  
ABJ25717  
ID ABJ25717 standard; Protein; 715 AA.  
XX  
XX AC ABJ25717;  
XX  
XX DT 16-APR-2003 (first entry)  
XX  
XX DE Aspergillus fumigatus essential gene protein #375.  
XX  
XX KW Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;  
XX cancer; contamination; biofilm; antibody; immune response.  
XX  
XX OS Aspergillus fumigatus.  
XX  
XX PN WO200286090-A2.  
XX  
XX PD 31-OCT-2002.  
XX  
XX PF 23-APR-2002; 2002WO-US13142.  
XX  
XX PR 23-APR-2001; 2001US-285697P.  
XX  
XX PR 27-APR-2001; 2001US-287066P.  
XX  
XX PR 05-JUN-2001; 2001US-295890P.  
XX  
XX PR 09-JUL-2001; 2001US-303899P.  
XX  
XX PR 31-AUG-2001; 2001US-316362P.  
XX  
XX PA (ELIT-) ELITRA PHARM INC.  
XX  
XX PI Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;  
XX WPI; 2003-093124/08.  
XX  
XX DR  
XX  
XX PT New purified or isolated nucleic acids of essential genes of  
XX Aspergillus fumigatus, useful for treating or preventing infections by  
XX A. fumigatus, or for treating a non-infectious disease in a subject  
XX e.g. cancer -  
XX  
XX PS Disclosure; Page -; 175pp; English.

CC The invention relates to novel purified or isolated nucleic acids of  
 CC essential genes of *Aspergillus fumigatus*. The isolated nucleic acids of  
 CC the invention are used to treat or prevent infections by a pathogenic  
 CC organism such as *A. fumigatus*, to treat a non-infectious disease in a  
 CC subject (e.g. cancer), to prevent or inhibit formation of an object  
 CC by *A. fumigatus*, or to prevent or inhibit formation on a surface of a  
 CC biofilm comprising *A. fumigatus*. The polynucleotides are useful for  
 CC expressing recombinant protein for characterisation, screening or  
 CC therapeutic use, as markers for host tissues in which the pathogenic  
 CC organisms invade or reside, for comparing with the DNA sequence of *A.*  
 CC *fumigatus* to identify duplicated genes or paralogues having the same or  
 CC similar biochemical activity and/or function, for comparing with DNA  
 CC sequences of other related or distant pathogenic organisms to identify  
 CC potential orthologous essential or virulence genes, for selecting and  
 CC making oligomers for attachment to a nucleic acid array for examination  
 CC of expression patterns, for raising anti-protein antibodies, as an  
 CC antigen to raise anti-DNA antibodies or to elicit another immune  
 CC response, and for identifying polynucleotides encoding the other protein  
 CC with which binding occurs or to identify inhibitors of the binding  
 CC interaction. The polypeptides may be used to raise antibodies or to  
 CC elicit immune response, as a reagent in assays designed to quantitatively  
 CC determine levels of the protein in biological fluids, as a marker for  
 CC host tissues in which pathogenic organism invade or reside, and to  
 CC isolate correlative receptors or ligands in the case of virulence  
 CC factors. This sequence represents a protein of one of the essential genes  
 CC of *Aspergillus fumigatus* of the invention.

XX Sequence 715 AA;  
 Query Match 88.2%; Score 30; DB 24; Length 715;  
 Best Local Similarity 85.7%; Pred. No. 4.1e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RRRGMAI 7  
 Db 473 RRRGMTI 479

RESULT 8  
 ABU26317  
 ID ABU26317 standard; Protein; 715 AA.  
 AC ABU26317;  
 DT 16-APR-2003 (first entry)  
 DE *Aspergillus fumigatus* essential gene protein #975.  
 KW Fungicide; cytostatic; essential gene; *Aspergillus fumigatus*; infection;  
 KW cancer; contamination; biofilm; antibody; immune response.  
 OS *Aspergillus fumigatus*.  
 XX WO200286090-A2.  
 XX 31-OCT-2002.  
 XX 23-APR-2002; 2002WO-US13142.  
 XX 23-APR-2001; 2001US-285697P.  
 XX 27-APR-2001; 2001US-287066P.  
 XX 05-JUN-2001; 2001US-295890P.  
 XX 09-JUL-2001; 2001US-303899P.  
 XX 31-AUG-2001; 2001US-316362P.  
 XX (ELIT-) ELITRA PHARM INC.  
 XX Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;  
 XX WPI; 2003-093124/08.

XX New purified or isolated nucleic acids of essential genes of  
 PT *Aspergillus fumigatus*, useful for treating or preventing infections by

PT *A. fumigatus*, or for treating a non-infectious disease in a subject  
 PT e.g. cancer -  
 PS Disclosure; Page -; 175pp; English.

XX The invention relates to novel purified or isolated nucleic acids of  
 CC essential genes of *Aspergillus fumigatus*. The isolated nucleic acids of  
 CC the invention are used to treat or prevent infections by a pathogenic  
 CC organism such as *A. fumigatus*, to treat a non-infectious disease in a  
 CC subject (e.g. cancer), to prevent or inhibit formation of an object  
 CC by *A. fumigatus*, or to prevent or inhibit formation on a surface of a  
 CC biofilm comprising *A. fumigatus*. The polynucleotides are useful for  
 CC expressing recombinant protein for characterisation, screening or  
 CC therapeutic use, as markers for host tissues in which the pathogenic  
 CC organisms invade or reside, for comparing with the DNA sequence of *A.*  
 CC *fumigatus* to identify duplicated genes or paralogues having the same or  
 CC similar biochemical activity and/or function, for comparing with DNA  
 CC sequences of other related or distant pathogenic organisms to identify  
 CC potential orthologous essential or virulence genes, for selecting and  
 CC making oligomers for attachment to a nucleic acid array for examination  
 CC of expression patterns, for raising anti-protein antibodies, as an  
 CC antigen to raise anti-DNA antibodies or to elicit another immune  
 CC response, and for identifying polynucleotides encoding the other protein  
 CC with which binding occurs or to identify inhibitors of the binding  
 CC interaction. The polypeptides may be used to raise antibodies or to  
 CC elicit immune response, as a reagent in assays designed to quantitatively  
 CC determine levels of the protein in biological fluids, as a marker for  
 CC host tissues in which pathogenic organism invade or reside, and to  
 CC isolate correlative receptors or ligands in the case of virulence  
 CC factors. This sequence represents a protein of one of the essential genes  
 CC of *Aspergillus fumigatus* of the invention.

XX Sequence 715 AA;

Query Match 88.2%; Score 30; DB 24; Length 715;  
 Best Local Similarity 85.7%; Pred. No. 4.1e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RRRGMAI 7  
 Db 473 RRRGMTI 479

RESULT 9  
 ABB71593  
 ID ABB71593 standard; Protein; 1335 AA.  
 AC ABB71593;  
 XX 26-MAR-2002 (first entry)  
 DT *Drosophila melanogaster* polypeptide SEQ ID NO 41571.  
 DE *Drosophila*; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 XX *Drosophila melanogaster*.  
 XX WO200171042-A2.  
 XX 27-SEP-2001.  
 XX 23-MAR-2001; 2001WO-US09231.  
 XX 23-MAR-2000; 2000US-191637P.  
 XX 11-JUL-2000; 2000US-0614150.  
 XX (PEKE ) PE CORP NY.  
 XX Venter JC, Adams M, Li PWD, Myers EW;  
 XX WPI; 2001-656860/75.  
 XX N-PSDB; ABU15696.

XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from *Drosophila* and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 XX Disclosure; SEQ ID NO 41571; 21pp + Sequence Listing; English.  
 XX  
 XX The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from *Drosophila*. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA  
 CC sequences (AB101840-AB116175) and the encoded proteins  
 CC (AB57737-AB872072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 XX Sequence 1335 AA;  
 SQ  
 Query Match 88.2%; Score 30; DB 22; Length 1335;  
 Best Local Similarity 71.4%; Pred. No. 7.4e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RRRGMWAI 7  
 Db :|||||:  
 999 RRRGMWAV 1005  
 XX  
 XX  
 XX RESULT 10  
 ABU00291  
 ID ABU00291 standard; Protein; 95 AA.  
 AC  
 AC ABU00291;  
 DT 17-JAN-2003 (first entry)  
 XX  
 XX Human novel polypeptide #384.  
 XX Human; genetic disorder; gene mapping; medical imaging; cancer;  
 KW neurodegenerative disorder; lymphoid cell disorder; osteoporosis;  
 KW Parkinson's disease; Alzheimer's disease; bone degenerative disorder;  
 KW osteoarthritis; periodontal disease; liver fibrosis; viral infection;  
 KW fungal infection; bacterial infection; autoimmune disease; diabetes;  
 KW atopic dermatitis.  
 XX  
 XX Homo sapiens.  
 OS  
 XX WO200274961-A1.  
 PN  
 XX 26-SEP-2002.  
 PD  
 XX 14-MAR-2002; 2002WO-US05109.  
 PF  
 XX 15-MAR-2001; 2001US-0810173.  
 PR  
 XX (HYSE-) HYSEQ INC.  
 PA  
 XX Tang YT, Zhou P, Goodrich R, Asundi V, Zhang J, Zhao QA, Ren F;  
 PI Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;  
 PI Weinman T, Wang J, Wang D, Drmanac RT;  
 XX WPI; 2003-040556/03.  
 DR N-PSDB; ABX05369.  
 DR  
 XX New isolated polypeptides and polynucleotides, useful for preventing,  
 PT treating or ameliorating medical conditions, such as cancer,  
 PT neurodegenerative disorders, lymphoid cell disorders, bone degenerative  
 PT disorders, and infections -  
 XX  
 XX Claim 9; SEQ ID NO 910; 235pp; English.  
 PS  
 XX

CC The invention relates to human polynucleotides and the polypeptides they  
 CC encode. The polynucleotides and polypeptides are useful in diagnostics,  
 CC forensics, gene mapping, medical imaging, identification of mutations  
 CC responsible for genetic disorders or other traits, assessing  
 CC biodiversity and producing many other types of data and products  
 CC dependent on DNA and amino acid sequences. They are also useful for  
 CC preventing, treating or ameliorating medical conditions, such as cancer,  
 CC neurodegenerative disorders (e.g. Parkinson's disease, Alzheimer's  
 CC disease), lymphoid cell disorders, osteoporosis, osteoarthritis, bone  
 CC degenerative disorders, periodontal disease, liver fibrosis, infections  
 CC (e.g. viral, fungal or bacterial) or autoimmune diseases (e.g. diabetes,  
 CC atopic dermatitis). Sequences ABG9888-ABG9989 and ABU00010-ABU00433  
 CC represent human polypeptides of the invention.  
 CC Note: The sequence data for this patent is not represented in the printed  
 CC specification but is based on sequence information supplied by the  
 CC European Patent Office.  
 XX  
 XX Sequence 95 AA;  
 SQ  
 Query Match 85.3%; Score 29; DB 24; Length 95;  
 Best Local Similarity 71.4%; Pred. No. 99;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RRRGMWAI 7  
 Db :|||||:  
 42 QRRGMWAV 48  
 XX  
 XX RESULT 11  
 AAY48598  
 ID AAY48598 standard; Protein; 129 AA.  
 AC  
 AC AAY48598;  
 XX  
 XX 08-DEC-1999 (first entry)  
 DT  
 XX Human breast tumour-associated protein 59.  
 DE  
 XX Expressed sequence tag; EST; human; breast; cancer; gene therapy;  
 KW treatment; tumour; cytostatic; medicament.  
 KW  
 KW Homo sapiens.  
 OS  
 XX DE19813839-A1.  
 PN  
 XX 23-SEP-1999.  
 PD  
 XX 20-MAR-1998; 98DE-1013839.  
 PF  
 XX 20-MAR-1998; 98DE-1013839.  
 PR  
 XX (META-) METAGEN GES GENOMFORSCHUNG MBH.  
 PA  
 XX Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, Rosentahl A;  
 PI WPI; 1999-528981/45.  
 DR N-PSDB; AAZ33663.  
 DR  
 XX Human nucleic acid sequences and protein products from tumor breast  
 PT tissue, useful for breast cancer therapy -  
 XX  
 XX Claim 22; 169; 188pp; German.  
 PS  
 XX This invention describes novel human nucleic acid sequences from tumor  
 CC breast tissue which have cytostatic activity. The nucleic acid sequences  
 CC can be used to produce and isolate full-length gene sequences. They can  
 CC be used to express proteins, which can be used as tools to find an  
 CC activity against breast cancer. The sequences can be used in sense or  
 CC antisense form. They are especially useful for medicaments for gene  
 CC therapy to treat breast cancer. AAY48540-Y48617 represent protein  
 CC fragments encoded by the expressed sequence tags described in the method  
 CC of the invention.  
 XX

SQ Sequence 129 AA;

Query Match 85.3%; Score 29; DB 20; Length 129;  
 Best Local Similarity 71.4%; Pred. No. 1.3e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRRGMAL 7  
 :|||||:  
 Db 98 QRRGMAL 104

RESULT 12

ABU05953

ID ABU05953 standard; Protein; 264 AA.

XX AC ABU05953;

XX DT 08-APR-2003 (first entry)

XX DE M. tuberculosis and M. leprae marker protein #604.

XX KW Mycobacterioses; survival; virulence; protective antigen; vaccine;  
 KW mycobacterial disease; tuberculosis; leprosy.

XX OS Mycobacterium tuberculosis.

OS Mycobacterium leprae.

XX XN W0200274903-A2.

XX PD 26-SEP-2002.

XX PF 22-FEB-2002; 2002WO-IB01973.

XX PR 22-FEB-2001; 2001US-270123P.

XX PA (INSP ) INST PASTEUR.

XX PI Cole S;

XX DR WPI; 2002-759885/82.

XX PT Identifying and selecting genes for survival or virulence of  
 PT mycobacteria by a comparative genomic analysis of the sequences of  
 PT Mycobacterium tuberculosis and M. leprae -

PS Claim 17; Page 821; 874pp; English.

XX This invention relates to a novel method for identifying essential genes  
 CC for survival or virulence of mycobacteria species. The method comprises  
 CC aligning the genomic sequence of a first mycobacterium species on a  
 CC genomic sequence of a second mycobacterium species and selecting a  
 CC polynucleotide sequence that is highly conserved in both genomes with no  
 CC counterparts in other bacterial genomic sequences and that corresponds  
 CC to an essential gene for the survival or virulence of mycobacterium  
 CC species. The method of the invention is useful for detecting M.  
 CC tuberculosis or M. leprae infection. The method reduces the number of  
 CC potential new targets and protective antigens for new drugs and vaccine  
 CC compositions to treat and prevent mycobacterial diseases, particularly  
 CC tuberculosis and leprosy. The present sequence represents a marker  
 CC protein from Mycobacterium tuberculosis and Mycobacterium leprae  
 CC identified using the method of the invention.

XX SQ Sequence 264 AA;

Query Match 85.3%; Score 29; DB 23; Length 264;  
 Best Local Similarity 71.4%; Pred. No. 2.6e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRRGMAL 7  
 :|||||:  
 Db 17 RRRGLAL 23

RESULT 13

AAW32427

ID AAW32427 standard; Protein; 267 AA.

XX AC AAW32427;

XX DT 08-JAN-1998 (first entry)

XX DE Mycobacterium tuberculosis antigen TbrA19.

XX KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;  
 KW skin testing; M.tuberculosis.

XX OS Mycobacterium tuberculosis.

XX PN W09709428-A2.

XX PD 13-MAR-1997.

XX PF 30-AUG-1996; 96WO-US14674.

XX PR 12-JUL-1996; 96US-0680574.

XX PR 01-SEP-1995; 95US-0523436.

XX PR 22-SEP-1995; 95US-0533634.

XX PR 22-MAR-1996; 96US-0620874.

XX PR 05-JUN-1996; 96US-0659683.

XX PA (CORI-) CORIXA CORP.

XX PI Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;  
 PI Twardzik DR, Vedvick TH;

XX DR WPI; 1997-192903/17.

XX DR N-PSDB; AAT91471.

XX PT New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are  
 PT useful in vaccines for prevention or treatment of tuberculosis, also  
 PT for diagnosis

XX PS Example 3; Page 103-104; 168pp; English.

XX A new immunogenic polypeptide has been developed comprising an  
 CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or  
 CC its variant differing only in conservative substitutions and/or  
 CC modifications). The present sequence represents a M.tuberculosis  
 CC antigen, TbrA19. The immunogenic protein, and fusion proteins  
 CC containing one or more of the proteins or one of the proteins plus  
 CC ESAT-6, are useful in vaccines, preferably when formulated with a  
 CC non-specific adjuvant, to induce an immune response against  
 CC M.tuberculosis (for treatment or prevention).

XX SQ Sequence 267 AA;

Query Match 85.3%; Score 29; DB 18; Length 267;  
 Best Local Similarity 71.4%; Pred. No. 2.6e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RRRGMAL 7  
 :|||||:  
 Db 20 RRRGLAL 26

RESULT 14

AAW32359

ID AAW32359 standard; Protein; 267 AA.

XX AC AAW32359;

XX DT 13-JAN-1998 (first entry)

XX DE Mycobacterium tuberculosis antigen TbrA19.

XX KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;

PR 11-OCT-1996; 96US-0730510.  
 XX (CORI-) CORIXA CORP.  
 XX Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;  
 PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;  
 XX WPI; 1998-261042/23.  
 DR N-PSDB; AAV64455.  
 XX Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used  
 PT to develop products for the detection of M. tuberculosis infection  
 PT and for diagnosis, treatment and prevention of tuberculosis  
 XX Example 3; Page 102-103; 230pp; English.  
 XX This sequence represents an immunogenic portion of a soluble  
 CC Mycobacterium tuberculosis (MT) antigen which can be used in a method  
 CC for inducing protective immunity against tuberculosis (TB). This sequence  
 CC can be formulated into vaccines and/or pharmaceutical compositions for  
 CC immunising against M. tuberculosis infection or may be used for the  
 CC diagnosis of tuberculosis.  
 XX SQ Sequence 267 AA;  
 Query Match 85.3%; Score 29; DB 19; Length 267;  
 Best Local Similarity 71.4%; Pred. No. 2.6e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RRRGNVAI 7  
 Db 20 RRRGLAL 26  
 Search completed: February 11, 2004, 17:02:47  
 Job time : 33.25 secs

XX skin testing; M.tuberculosis.  
 XX Mycobacterium tuberculosis.  
 XX WO9709429-A2.  
 XX 13-MAR-1997.  
 XX 30-AUG-1996; 96WO-US14675.  
 XX 12-JUL-1996; 96US-0680573.  
 PR 01-SEP-1995; 95US-0523435.  
 PR 22-SEP-1995; 95US-0532136.  
 PR 22-MAR-1996; 96US-0620280.  
 PR 05-JUN-1996; 96US-0658800.  
 XX (CORI-) CORIXA CORP.  
 XX Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;  
 PI Twardzik DR, Vedvick TH;  
 XX WPI; 1997-192904/17.  
 DR N-PSDB; AAT91408.  
 XX New immunogenic polypeptide(s) from soluble M. tuberculosis antigens  
 PT - useful for diagnosis of M. tuberculosis infection  
 XX Example 3; Page 111-112; 190pp; English.  
 XX A new immunogenic polypeptide has been developed comprising an  
 CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or  
 CC its variant differing only in conservative substitutions and/or  
 CC modifications). The present sequence represents a M.tuberculosis  
 CC antigen, Tbra19. The immunogenic polypeptide can be used to diagnose  
 CC M.tuberculosis infection by forming complexes with specific  
 CC antibodies in the sample. Fragments of DNA encoding the immunogenic  
 CC polypeptide can be used as diagnostic primers or probes and agents  
 CC that bind to the antigen, especially monoclonal antibodies or  
 CC equivalent polyclonal antibodies, are also used for diagnosis.  
 XX SQ Sequence 267 AA;  
 Query Match 85.3%; Score 29; DB 18; Length 267;  
 Best Local Similarity 71.4%; Pred. No. 2.6e+02;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RRRGNVAI 7  
 Db 20 RRRGLAL 26  
 RESULT 15  
 AAW81662  
 ID AAW81662 standard; Protein; 267 AA.  
 XX AAW81662;  
 XX 27-JAN-1999 (first entry)  
 XX M. tuberculosis immunogenic polypeptide Tbra19.  
 XX Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;  
 KW vaccine; pharmaceutical; infection; diagnosis.  
 XX Mycobacterium tuberculosis.  
 OS WO9816646-A2.  
 PN 23-APR-1998.  
 XX 07-OCT-1997; 97WO-US18293.  
 XX 13-MAR-1997; 97US-0818112.



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 16:52:34 ; Search time 24.5833 seconds  
(without alignments)  
73.479 Million cell updates/sec

Title: US-09-901-187C-6  
Perfect score: 41  
Sequence: 1 TKHGRK 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL.23.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phage.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_virus.\*  
16: sp\_bacteriap.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	36	87.8	419	16 Q9K9W7	Q9K9W7 bacillus ha
2	36	87.8	656	5 Q9SP08	Q9SP08 scutigera c
3	34	82.9	612	16 Q9PF38	Q9PF38 xylella fas
4	33	80.5	82	10 Q9LH73	Q9LH73 oryza sativ
5	33	80.5	266	12 Q55335	Q55335 canine dist
6	33	80.5	266	12 Q55336	Q55336 canine dist
7	33	80.5	266	12 Q55333	Q55333 canine dist
8	33	80.5	304	4 Q8TEU2	Q8TEU2 homo sapien
9	33	80.5	304	4 Q9V4X2	Q9V4X2 homo sapien
10	33	80.5	314	4 Q9BPW1	Q9BPW1 homo sapien
11	33	80.5	315	6 Q9SG99	Q9SG99 bos taurus
12	33	80.5	369	4 Q9NSU2	Q9NSU2 homo sapien
13	33	80.5	435	2 Q8XRU8	Q8XRU8 uncultured
14	33	80.5	443	4 Q96MB1	Q96MB1 homo sapien
15	33	80.5	483	16 Q9ACZ3	Q9ACZ3 streptomyce
16	33	80.5	547	4 Q60364	Q60364 homo sapien

17	33	80.5	555	5 Q20673	Q20673 caenorhabdi
18	33	80.5	567	16 Q8YK20	Q8YK20 anabaena sp
19	33	80.5	573	5 Q9YIX9	Q9YIX9 ephydataia f
20	33	80.5	597	16 Q8IOW2	Q8IOW2 anabaena sp
21	33	80.5	677	4 Q9P242	Q9P242 homo sapien
22	33	80.5	743	4 Q9HOK2	Q9HOK2 homo sapien
23	33	80.5	774	4 Q9BXH9	Q9BXH9 homo sapien
24	33	80.5	774	4 Q96IC2	Q96IC2 homo sapien
25	33	80.5	1179	5 Q9NCQ0	Q9NCQ0 aedes aegypt
26	33	80.5	1698	2 Q9LC00	Q9LC00 staphylococ
27	32	78.0	104	5 Q9GF72	Q9GF72 drosophila
28	32	78.0	104	5 Q9GF75	Q9GF75 drosophila
29	32	78.0	104	5 Q9GN84	Q9GN84 drosophila
30	32	78.0	104	5 Q9GN82	Q9GN82 drosophila
31	32	78.0	104	5 Q9GN83	Q9GN83 drosophila
32	32	78.0	105	5 Q9GF79	Q9GF79 drosophila
33	32	78.0	105	5 Q9GF78	Q9GF78 drosophila
34	32	78.0	105	5 Q9GNH2	Q9GNH2 drosophila
35	32	78.0	106	5 Q9GF80	Q9GF80 drosophila
36	32	78.0	311	16 Q9PAE5	Q9PAE5 xylella fas
37	32	78.0	405	2 Q8GRC1	Q8GRC1 deltia sp.
38	32	78.0	436	16 Q9KVG7	Q9KVG7 vibrio chol
39	32	78.0	610	11 Q8EVT0	Q8EVT0 mus musculu
40	32	78.0	698	5 Q9GRX4	Q9GRX4 drosophila
41	32	78.0	894	10 Q84P5	Q84P5 zea mays (m
42	32	78.0	1022	5 Q96OV3	Q96OV3 drosophila
43	32	78.0	1308	4 Q96IF2	Q96IF2 homo sapien
44	32	78.0	1738	10 Q9SI41	Q9SI41 arabidopsis
45	32	78.0	1811	4 Q9H7S0	Q9H7S0 homo sapien

## ALIGNMENTS

### RESULT 1

Q9K9W7 PRELIMINARY; PRT; 419 AA.  
ID Q9K9W7  
AC Q9K9W7  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
DE Hypothetical protein BH2528.  
GN BH2528.  
OS Bacillus halodurans.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_taxid=86665;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C-125 / JCM 9153;  
RX MEDLINE=20512582; PubMed=11058132;  
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,  
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,  
RA Horikoshi K.;  
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
RT halodurans and genomic sequence comparison with Bacillus subtilis.";  
RL Nucleic Acids Res. 28:4317-4331(2000).  
DR EMBL; AP001515; BAB06247.1; -  
DR InterPro; IPR007114; MPS.  
DR InterPro; IPR005828; Sub transporter.  
DR Pfam; PF00083; sugar tr.; 1.  
DR PROSITE; PS50850; MES; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 419 AA; 45104 MW; 5ABD7E18283489B6 CRC64;

Query Match 87.8%; Score 36; DB 16; Length 419;

Best Local Similarity 100.0%; Pred. No. 16;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 KHGRK 7

DB 69 KHGRK 74

## RESULT 2

Q95F08 PRELIMINARY; PRT; 656 AA.  
 AC Q95F08;  
 DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)  
 DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)  
 DE Hemocyanin subunit 1 precursor.  
 GN HCL  
 OS Scutigera coleoptrata (house centipede)  
 OC Eukaryota; Metazoa; Arthropoda; Myriapoda; Chilopoda; Notostigmophora;  
 OC Scutigermorpha; Scutigeridae; Scutigera.  
 NCBI\_TaxID=29022;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Kusche K., Burmester T.;  
 RT "Molecular characterization of the hemocyanin from the centipede,  
 RL Scutigera coleoptrata";  
 DR EMBL; AJ344359; CAC69246.1; -;  
 DR InterPro; IPR000896; Hemocyanin.  
 DR InterPro; IPR005203; hemocyanin C.  
 DR InterPro; IPR005204; hemocyanin N.  
 DR InterPro; IPR000169; SHprot acsite.  
 DR InterPro; IPR002227; Tyrosinase.  
 DR Pfam; PF00372; hemocyanin; 1.  
 DR Pfam; PF03723; hemocyanin; C; 1.  
 DR Pfam; PF03722; hemocyanin; N; 1.  
 DR PROSITE; PS00209; HEMOCYANIN; 1.  
 DR PROSITE; PS00210; HEMOCYANIN; 2; 1.  
 DR PROSITE; PS00639; THIOLE PROTEASE; HIS; 1.  
 DR PROSITE; PS00498; TYROSINASE; 2; 1.  
 KW Signal.  
 FT SIGNAL.  
 FT CHAIN.  
 SQ SEQUENCE 656 AA; 76382 MW; 78F06FE39140AE07 CRC64;

Query Match 87.8%; Score 36; DB 5; Length 656;  
 Best Local Similarity 100.0%; Pred. No. 25;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

## QY 2 KHGPRK 7

Db 211 KHGPRK 216

## RESULT 3

Q9PF38 PRELIMINARY; PRT; 612 AA.  
 AC Q9PF38;  
 DT 01-OCT-2000 (TREMELrel. 15, Created)  
 DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)  
 DT 01-OCT-2002 (TREMELrel. 22, Last annotation update)  
 DE Beta-galactosidase (EC 3.2.1.23) (lactase).  
 GN Xf0840.  
 OS Xylella fastidiosa.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;  
 OC Xanthomonadaceae; Xylella.  
 NCBI\_TaxID=2371;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=9a5C;  
 RX MEDLINE=20365717; PubMed=10910347;  
 RA Simpson A.J.G., Reinach P.C., Arruda P., Abreu F.A., Acencio M.,  
 RA Alvaranga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,  
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,  
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carter H.,  
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,  
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,  
 RA Pacinani A.P., Ferreira A.J.S., Ferreira V.C.A., Fetto J.A.,  
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,  
 RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,  
 RA Ho P.L., Hobeisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,

RA Krieger J.E., Kuranae E.E., Laigret F., Lambais M.R., Leite L.C.C.,  
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,  
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.B.,  
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,  
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,  
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,  
 RA Nhani A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,  
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,  
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,  
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,  
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,  
 RA da Silva J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,  
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,  
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,  
 RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;  
 RT "The genome sequence of the plant pathogen Xylella fastidiosa";  
 CC Nature 406:151-159(2000).  
 CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL, NON-REDUCING BETA-D-  
 CC GALACTOSE RESIDUES IN BETA-D-GALACTOSIDES.  
 CC EMBL; AB003923; AAF83650.1; -;  
 DR InterPro; IPR001944; Glyco\_hydro\_35.  
 DR Pfam; PF01301; Glyco\_hydro\_35; 1.  
 DR PRINTS; PR00742; GLYCOLASE35.  
 DR PROSITE; PS01182; GLYCOSYL HYDROL F35; 1.  
 KW Glycosidase; Hydrolase; Complete Proteome.  
 SQ SEQUENCE 612 AA; 67893 MW; B894D6A3EB7895CB CRC64;

Query Match 82.9%; Score 34; DB 16; Length 612;  
 Best Local Similarity 85.7%; Pred. No. 63;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

## QY 1 TKHGRK 7

Db 413 TKHGRK 419

## RESULT 4

Q8LHY3 PRELIMINARY; PRT; 82 AA.  
 AC Q8LHY3;  
 DT 01-OCT-2002 (TREMELrel. 22, Created)  
 DT 01-OCT-2002 (TREMELrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)  
 DE P0022B05.17 protein.  
 GN P0022B05.17.  
 OS Oryza sativa (japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzoideae; Oryza.  
 NCBI\_TaxID=39947;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=cv. Nipponbare;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, PAC  
 RT clone:P0022B05.17";  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AP004262; BAC10813.1; -;  
 DR Gramene; Q8LHY3; -;  
 SQ SEQUENCE 82 AA; 10078 MW; 2D6FFE011F7BDE39 CRC64;

Query Match 80.5%; Score 33; DB 10; Length 82;  
 Best Local Similarity 83.3%; Pred. No. 13;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

## QY 2 KHGPRK 7

Db 20 KHGPRK 25

## RESULT 5

```

OS5335
ID OS5335 PRELIMINARY; PRT; 266 AA.
AC OS5335;
DT 01-JUN-1998 (TRENBLrel. 06, Created)
DT 01-JUN-1998 (TRENBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE Fusion protein (Fragment).
GN F.
OS Canine distemper virus.
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Paramyxoviridae; Paramyxovirinae; Morbillivirus.
OX NCBI_TaxID=11232;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Dog #5804/89;
RA Liemann H., Harder T., Haas L.;
RT "Genetic analysis of the central untranslated genome region and the
RT proximal coding part of the F gene of wild-type and vaccine distemper
RT morbilliviruses.";
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF026241; AAB88266.1; -
DR InterPro; IPR000776; Fusion_gly.
DR Pfam; PF00523; fusion_gly; 1.
FT NON_TER 266
SQ SEQUENCE 266 AA; 29191 MW; 7A80B29F09B972EE CRC64;

Query Match 80.5%; Score 33; DB 12; Length 266;
Best Local Similarity 83.3%; Pred. No. 44;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TKHGR 6
Db 46 TRHGR 51

RESULT 6
ID OS5336 PRELIMINARY; PRT; 266 AA.
AC OS5336;
DT 01-JUN-1998 (TRENBLrel. 06, Created)
DT 01-JUN-1998 (TRENBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE Fusion protein (Fragment).
GN F.
OS Canine distemper virus.
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Paramyxoviridae; Paramyxovirinae; Morbillivirus.
OX NCBI_TaxID=11232;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Dog #9999/92;
RA Liemann H., Harder T., Haas L.;
RT "Genetic analysis of the central untranslated genome region and the
RT proximal coding part of the F gene of wild-type and vaccine distemper
RT morbilliviruses.";
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF026242; AAB88267.1; -
DR InterPro; IPR000776; Fusion_gly.
DR Pfam; PF00523; fusion_gly; 1.
FT NON_TER 266
SQ SEQUENCE 266 AA; 29135 MW; 16A6FEA46AEC1DAC CRC64;

Query Match 80.5%; Score 33; DB 12; Length 266;
Best Local Similarity 83.3%; Pred. No. 44;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TKHGR 6
Db 46 TRHGR 51

RESULT 7
OS5333

```

```

ID OS5333 PRELIMINARY; PRT; 266 AA.
AC OS5333;
DT 01-JUN-1998 (TRENBLrel. 06, Created)
DT 01-JUN-1998 (TRENBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE Fusion protein (Fragment).
GN F.
OS Canine distemper virus.
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Paramyxoviridae; Paramyxovirinae; Morbillivirus.
OX NCBI_TaxID=11232;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Dog #2544/95;
RA Liemann H., Harder T., Haas L.;
RT "Genetic analysis of the central untranslated genome region and the
RT proximal coding part of the F gene of wild-type and vaccine distemper
RT morbilliviruses.";
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF026239; AAB88264.1; -
DR InterPro; IPR000776; Fusion_gly.
DR Pfam; PF00523; fusion_gly; 1.
FT NON_TER 266
SQ SEQUENCE 266 AA; 29164 MW; CBC813FC0A366375 CRC64;

Query Match 80.5%; Score 33; DB 12; Length 266;
Best Local Similarity 83.3%; Pred. No. 44;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TKHGR 6
Db 46 TRHGR 51

RESULT 8
ID OS7E2 PRELIMINARY; PRT; 304 AA.
AC OS7E2;
DT 01-JUN-2002 (TRENBLrel. 21, Created)
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Three prime repair exonuclease 1.
GN TREX1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Rieder M.J., Braun A.C., Montoya M.A., Chung M.-W., Nguyen C.P.,
RA Nguyen D.A., Livingston R.J., Poel C.L., Robertson P.D.,
RA Schackwitz W.S., Sherwood J.K., Witrak L.A., Nickerson D.A.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF483777; AAL82504.1; -
DR InterPro; IPR006055; Exonuclease.
DR Pfam; PF00929; Exonuclease; 1.
KW Exonuclease.
SQ SEQUENCE 304 AA; 32276 MW; 922048DCC4122124 CRC64;

Query Match 80.5%; Score 33; DB 4; Length 304;
Best Local Similarity 71.4%; Pred. No. 51;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 TKHGR 7
Db 159 SEHGR 165

RESULT 9
ID Q9V4X2 PRELIMINARY; PRT; 304 AA.
ID Q9V4X2;
AC Q9V4X2;
DT 01-NOV-1999 (TRENBLrel. 12, Created)

```

DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE Dnaase III.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9321773; PubMed=10393201;  
 RA Hoss M., Robins P., Naven T.J., Pappin D.J., Sgouros J., Lindahl T.;  
 RT "A human DNA editing enzyme homologous to the Escherichia coli  
 RT DnaQ/MutD protein.";  
 RL EMBO J. 18:3868-3875(1999).  
 DR EMBL; AJ243797; CAB50866.1; -.  
 DR InterPro; IPR006055; Exonuclease.  
 DR Pfam; PF00929; Exonuclease; 1.  
 KW Exonuclease.  
 SQ SEQUENCE 304 AA; 32375 MW; 923AF8DCDEA22124 CRC64;  
 Query Match 80.5%; Score 33; DB 4; Length 304;  
 Best Local Similarity 71.4%; Pred. No. 51;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 TKHGPRK 7  
 Db 159 SEHGPRK 165  
 RESULT 10  
 Q9BFW1  
 ID Q9BFW1 PRELIMINARY; PRT; 314 AA.  
 AC Q9BFW1;  
 DT 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DE 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE 3'-5' exonuclease TREX1.  
 GN TREX1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99321890; PubMed=10391904;  
 RA Mazur D.J., Perrino F.W.;  
 RT "Identification and expression of the TREX1 and TREX2 cDNA sequences  
 RT encoding mammalian 3'--5' exonucleases.";  
 RL J. Biol. Chem. 274:19655-19660(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX Mazur D.J., Perrino F.W.;  
 RT "Structure and expression of the TREX1 and TREX2 3'-5' exonuclease  
 RT genes.";  
 RL J. Biol. Chem. 274:19655-19660(1999).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX Mazur D.J., Mazur D.J.;  
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF319569; AAK07616.1; -.  
 DR EMBL; AF319566; AAK07613.1; -.  
 DR EMBL; AF319567; AAK07614.1; -.  
 DR EMBL; AF151105; AAD49774.2; -.  
 DR InterPro; IPR006055; Exonuclease.  
 DR Pfam; PF00929; Exonuclease; 1.  
 KW Exonuclease.  
 SQ SEQUENCE 314 AA; 33212 MW; EE8F63B6496D72F4 CRC64;  
 Query Match 80.5%; Score 33; DB 4; Length 314;  
 Best Local Similarity 71.4%; Pred. No. 52;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 TKHGPRK 7

Db 169 SEHGPRK 175  
 RESULT 11  
 Q9BG99  
 ID Q9BG99 PRELIMINARY; PRT; 315 AA.  
 AC Q9BG99;  
 DT 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE 3'-5' exonuclease TREX1.  
 GN TREX1.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99321890; PubMed=10391904;  
 RA Mazur D.J., Perrino F.W.;  
 RT "Identification and expression of the TREX1 and TREX2 cDNA sequences  
 RT encoding mammalian 3'--5' exonucleases.";  
 RL J. Biol. Chem. 274:19655-19660(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX Mazur D.J., Perrino F.W.;  
 RT "Structure and expression of the TREX1 and TREX2 3'-5' exonuclease  
 RT genes.";  
 RL J. Biol. Chem. 274:19655-19660(1999).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX Mazur D.J., Mazur D.J.;  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF319575; AAK07622.1; -.  
 KW Exonuclease.  
 SQ SEQUENCE 315 AA; 33132 MW; 90AD66D1513DFDE6 CRC64;  
 Query Match 80.5%; Score 33; DB 6; Length 315;  
 Best Local Similarity 71.4%; Pred. No. 52;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 TKHGPRK 7  
 Db 169 SEHGPRK 175  
 RESULT 12  
 Q9NSU2  
 ID Q9NSU2 PRELIMINARY; PRT; 369 AA.  
 AC Q9NSU2;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE Hypothetical protein (3'-5' exonuclease TREX1-like protein) (Three  
 DE prime repair exonuclease 1).  
 GN DKFZP434J0310 OR TREX1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX TissueThesit;  
 RC Dueterhoest A., Lauber J., Mewes H.W., Gassenhuber J., Wiemann S.;  
 RA Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99321890; PubMed=10391904;  
 RA Mazur D.J., Perrino F.W.;  
 RT "Identification and expression of the TREX1 and TREX2 cDNA sequences  
 RT encoding mammalian 3'--5' exonucleases.";  
 RL J. Biol. Chem. 274:19655-19660(1999).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX Mazur D.J., Perrino F.W.;

```

RT "Structure and expression of the TREX1 and TREX2 3'-5' exonuclease
RL genes"; Chem. 0:0-0(2001).
RN J. Biol.
RP [4]
RC SEQUENCE FROM N.A.
RA TISSUE=Eye;
RA Strausberg R.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL137745; CAB70902.1; -
DR EMBL; AF319568; AAK07615.1; -
DR EMBL; BC023630; AAH23630.1; -
DR InterPro; IPR006055; Exonuclease.
DR Pfam; PF00929; Exonuclease; 1.
KW Hypothetical protein; Exonuclease.
SQ SEQUENCE 369 AA; 38922 MW; 42B79047A9AD9837 CRC64;

Query Match 80.5%; Score 33; DB 4; Length 369;
Best Local Similarity 71.4%; Pred. No. 62;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 TKHGPRK 7
DB 224 SEHGPRK 230

RESULT 13
Q8KYU8 PRELIMINARY; PRT; 435 AA.
AC Q8KYU8;
DT 01-OCT-2002 (TREMBLrel. 22, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Heat shock protein HslVU, ATPase subunit HslU.
GN HSLU.
OS uncultured proteobacterium.
OC Bacteria; Proteobacteria; environmental samples.
CX NCBI_TaxID=153809;
RN [1]
RL SEQUENCE FROM N.A.
RX MEDLINE=21822632; PubMed=11832943;
RA Beja O., Suzuki M.T., Heidelberg J.F., Nelson W.C., Preston C.M.,
RA Hamada T., Eisen J.A., Fraser C.M., DeLong E.F.;
RT "Unsuspected diversity among marine aerobic anoxygenic phototrophs.";
RL Nature 415:630-633(2002).
DR EMBL; AB008921; BAM48725.1; -
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003959; AAA_ATPase_cent.
DR Pfam; PF00004; AAA; 1.
DR SMART; SM00382; AAA; 1.
DR TIGRFAMs; TIGR00390; hslU; 1.
KW ATP-binding.
SQ SEQUENCE 435 AA; 48117 MW; 4C758FC1738748DA CRC64;

Query Match 80.5%; Score 33; DB 2; Length 435;
Best Local Similarity 85.7%; Pred. No. 73;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TKHGPRK 7
DB 284 TRHGPRK 290

RESULT 14
Q96MB1 PRELIMINARY; PRT; 443 AA.
AC Q96MB1;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Hypothetical protein FLJ32692.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=Testis;
RA Ishibashi T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y.,
RA Hotuta T., Hiraoaka S., Murakawa K., Takiguchi S., Kusano J.,
RA Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,
RA Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S.,
RA Yamamoto J., Isono Y., Kawai-Hio Y., Saiko K., Nishikawa T.,
RA Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Nagatsuma M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA Suzuki Y., Sugano S., Negahari K., Masuo Y., Nagai K., Isogai T.;
RL "NDO human cDNA sequencing project.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK057254; BAB71397.1; -
DR InterPro; IPR006055; Exonuclease.
DR Pfam; PF00929; Exonuclease; 1.
DR SMART; SM00479; EXOIII; 1.
KW Hypothetical protein.
SQ SEQUENCE 443 AA; 49912 MW; B8E469BC931C5318 CRC64;

Query Match 80.5%; Score 33; DB 4; Length 443;
Best Local Similarity 83.3%; Pred. No. 75;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 KHGPRK 7
DB 381 KHGPRK 386

RESULT 15
Q9ACZ3 PRELIMINARY; PRT; 483 AA.
AC Q9ACZ3;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Hypothetical protein SCP1.158.
GN SCP1.158.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycetaceae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra J., Hornsby T., Howarth S.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Oliver S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL590463; CAC36679.1; -
KW Hypothetical protein; Plasmid; Complete proteome.
SQ SEQUENCE 483 AA; 53756 MW; 3B1787A0D586C7E2 CRC64;

Query Match 80.5%; Score 33; DB 16; Length 483;
Best Local Similarity 83.3%; Pred. No. 82;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TKHGPR 6
DB 368 TRHGPR 373

```

us-09-901-187c-6.rspt

Thu Feb 12 09:07:47 2004

Search completed: February 11, 2004, 17:09:28  
Job time : 26.5833 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 16:51:43 ; Search time 5.16667 Seconds  
(without alignments)  
63.714 Million cell updates/sec

Title: US-09-901-187C-6

Perfect score: 41

Sequence: 1 TKHGPRK 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	80.5	1637	1 MRSP_STAAU	P80544 staphylococ
2	32	78.0	1036	1 SMO_DROME	P91682 drosophila
3	32	78.0	1701	1 US20_HUMAN	O75643 homo sapien
4	31	75.6	205	1 EFAL_HUMAN	P20827 homo sapien
5	31	75.6	205	1 EFAL_MOUSE	P52793 mus muscucu
6	31	75.6	205	1 EFAL_RAT	P97553 rattus norv
7	31	75.6	245	1 YG24_YEAST	P53237 saccharomyc
8	31	75.6	315	1 MRB6_BACSP	P43420 bacillus sp
9	31	75.6	445	1 P021_PORCA	Q03273 popillia ja
10	31	75.6	447	1 LSC_FARPC	P83304 parkia plat
11	31	75.6	751	1 FPI_MYTGA	Q27409 mytilus gal
12	31	75.6	807	1 MCM3_XENLA	P49739 xenopus lae
13	31	75.6	1011	1 POLG_FMDVT	P15072 foot-and-mo
14	31	75.6	2332	1 POLG_FMDVA	P03308 f genome po
15	31	75.6	2332	1 POLG_FMDVO	P03305 f genome po
16	31	75.6	2333	1 POLG_FMDVI	P03306 f genome po
17	31	75.6	2336	1 POLG_FMDVZ	P49303 f genome po
18	31	75.6	3511	1 MY15_MOUSE	Q94224 mus muscucu
19	30	73.2	88	1 YXCB_LACLA	Q9CFO9 lactococcus
20	30	73.2	103	1 VC20_VACCC	P21104 vaccinia vi
21	30	73.2	211	1 RL4_LEPIN	Q9X435 leptospira
22	30	73.2	279	1 Y939_METKA	P58851 methanopyru
23	30	73.2	303	1 YLYB_BACSU	Q45480 bacillus su
24	30	73.2	324	1 LX82_PHOLU	P23147 photorhabdu
25	30	73.2	356	1 RC6F_ECOLI	P03016 escherichia
26	30	73.2	356	1 RC6F_SALTI	Q82294 salmonella
27	30	73.2	356	1 RECF_SALTY	P24900 salmonella
28	30	73.2	1178	1 RPOC_CLOPE	Q93R87 clostridium
29	30	73.2	1182	1 RPOC_CLOAB	Q97E80 clostridium
30	29	70.7	117	1 RL13_ECOLI	P02369 escherichia
31	29	70.7	117	1 RL13_SALTY	Q82Lm1 salmonella
32	29	70.7	118	1 RL13_SHEVI	Q980r1 shewanella
33	29	70.7	118	1 RL13_XANCP	Q923f0 xanthomonas

#### ALIGNMENTS

##### RESULT 1

ID	MRSP_STAAU	STANDARD;	PRT;	1637 AA.
AC	P80544; Q9ZF62;			
DT	01-FEB-1996 (Rel. 33, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Methicillin-resistant surface protein precursor.			
GN	PLS.			
OS	Staphylococcus aureus.			
OC	Bacteria; Firmicutes; Bacillales; Staphylococcus.			
OX	NCBI_TaxID=1280;			
RN	(1)			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Isolate 1061;			
RA	Savolainen K., Kuusela P., Paulin L., Korhonen T.K.;			
RT	"Pls, a large repeat-rich surface protein of methicillin resistant			
RT	Staphylococcus aureus.";			
RL	Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.			
RN	(2)			
RP	SEQUENCE OF 624-628; 676-682; 938-948; 1156-1168; 1176-1185;			
RP	1199-1205 AND 1217-1224.			
RC	STRAIN=Isolate 1061;			
RX	MEDLINE=96270743; Pubmed=8665912;			
RA	Hilden P., Savolainen K., Tynnelae J., Vuento M., Kuusela P.;			
RT	"Purification and characterisation of a plasmin-sensitive surface			
RT	protein of Staphylococcus aureus.";			
RL	Eur. J. Biochem. 236:904-910(1996).			
CC	-!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by			
CC	an amide bond (Potential).			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
DR	EMBL; AF115379; AAC09131.1; ...			
DR	InterPro; IPR005877; Gpos_Y5IRK.			
DR	InterPro; IPR001899; Gram_pos_anchor.			
DR	InterPro; IPR006192; LPXTG.			
DR	Pfam; PF00745; Gram_pos_anchor; 1.			
DR	TIGRFAMs; TIGR01167; LPXTG signal; 1.			
DR	TIGRFAMs; TIGR01168; Y5IRK signal; 1.			
DR	PROSITE; PS50847; GRAM_POS_ANCHORING; 1.			
KW	Antibiotic resistance; Glycoprotein; Cell wall; Peptidoglycan-anchor;			
	Repeat; Signal.			
FT	SIGNAL	48		POTENTIAL.
FT	CHAIN	49	1601	METHICILLIN-RESISTANT SURFACE PROTEIN.
FT	PROPEP	1602	1637	REMOVED BY SORTASE (POTENTIAL).
FT	DOMAIN	1301	1582	141 X 2 AA TANDEM REPEATS OF D-(SAG).
FT	SITE	1598	1602	LPXTG SORTING SIGNAL (POTENTIAL).
FT	MOD_RES	1601	1601	AMIDE-LINKED TO CELL WALL (POTENTIAL).

34 29 70.7 123 1 RS13\_MYCGA Q9rdv8 mycoplasma  
35 29 70.7 123 1 RS13\_MYCGE P47421 mycoplasma  
36 29 70.7 124 1 RS13\_MYCPN Q50297 mycoplasma  
37 29 70.7 125 1 RS13\_THERH P80377 thermus the  
38 29 70.7 172 1 YM19\_MARPO P38462 marchantia  
39 29 70.7 207 1 RR4\_CHLAVU P32975 chlorella v  
40 29 70.7 612 1 GAG\_JSRV P31622 sheep pulmo  
41 29 70.7 624 1 MYB\_XENLA Q08759 xenopus lae  
42 29 70.7 808 1 SEAZ\_MYCTU Q50612 mycobacteri  
43 29 70.7 837 1 ROD1\_YEAST Q02805 saccharomyc  
44 29 70.7 1493 1 NEO1\_MOUSE P97798 mus musculu  
45 29 70.7 1932 1 PAB1\_SCHPO O59722 schizosacch

SQ SEQUENCE 1637 AA; 174573 MW; 75BE9ADB469BD309 CRC64;

Query Match 80.5%; Score 33; DB 1; Length 1637;  
 Best Local Similarity 83.3%; Pred. No. 85;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TKHGPR 6  
 |||||  
 Db 781 TKHGPK 786

RESULT 2

SMO\_DROME STANDARD; PRT; 1036 AA.

AC P91682; Q9VPM8;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Smoothed protein precursor (dSMO) (SMO) (Smooth).  
 GN SMO OR CG11561.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydriidae; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96319725; PubMed=8706127;  
 RA Alcedo J., Avizonon M., von Ohlen T., Noll M., Hooper J.E.;  
 RT "The drosophila smoothed gene encodes a seven-pass membrane  
 protein, a putative receptor for the hedgehog signal."  
 RL Cell 86:221-232 (1996).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96320560; PubMed=8700230;  
 RA van den Heuvel M., Ingham P.W.;  
 RT "smoothed encodes a receptor-like serpentine protein required for  
 hedgehog signalling."  
 RL Nature 382:547-551 (1996).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkeley.  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.P.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer V.G., Champagne M., Pfeiffer B.D.,  
 RA Van K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Fodor C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.R., Gu Z., Guan P., Harris M.,  
 RA Harris N.H., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush C., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.B., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattai B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.B.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 RA Palazzolo M., Pattan K.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington G., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195 (2000).  
 CC -!- FUNCTION: SEGMENT POLARITY PROTEIN REQUIRED FOR CORRECT PATTERNING  
 CC OF EVERY SEGMENT. G PROTEIN-COUPLED RECEPTOR THAT ASSOCIATES WITH  
 CC THE PATCHED PROTEIN (PTC) TO TRANSDUCE THE HEDGEHOG (HH) SIGNAL  
 CC THROUGH THE ACTIVATION OF AN INHIBITORY G-PROTEIN. IN THE ABSENCE  
 CC OF HH, PTC REPASSES THE CONSTITUTIVE SIGNALING ACTIVITY OF SMO  
 CC THROUGH FUSED (FU).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- DEVELOPMENTAL STAGE: EXPRESSED AT ALL DEVELOPMENTAL STAGES, THOUGH  
 CC THE LEVELS VARY.  
 CC -!- SIMILARITY: BELONGS TO FAMILY FZ/SMO OF G-PROTEIN COUPLED  
 CC RECEPTORS.  
 CC -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1, MET-9, MET-13 OR MET-14 IS  
 CC THE INITIATOR.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
 CC EMBL; U87613; AAC33180.1; -;  
 CC EMBL; AF030334; AAB84275.1; -;  
 CC EMBL; AE003590; AAP51518.2; -;  
 CC FlyBase; FBGN003444; smc.  
 CC GO; GO:0007350; P:blastoderm segmentation; IMP.  
 CC GO; GO:0007455; P:eye-antennal disc metamorphosis; IGI.  
 CC GO; GO:0007346; P:regulation of mitotic cell cycle; IMP.  
 CC InterPro; IPR000539; FzIzled.  
 CC InterPro; IPR000024; Fz\_domain.  
 CC InterPro; IPR000832; GPCR\_secretin.  
 CC Pfam; PF01534; FzIzled; 1.  
 CC Pfam; PF01392; Fz; 1.  
 CC PRINTS; PRO0489; FRIZZLED.  
 CC SMART; SM00063; FRI; 1.  
 CC PROSITE; PS50038; FZ; 1.  
 CC PROSITE; PS50261; G\_PROTEIN\_RECEP\_F2\_4; 1.  
 CC G-protein coupled receptor; transmembrane; Glycoprotein; Signal;  
 CC Developmental protein.  
 CC SIGNAL 1 31  
 CC CHAIN 32 1036  
 CC DOMAIN 32 258  
 CC EXTRACELLULAR (POTENTIAL).  
 CC TRANSMEM 259 279  
 CC 1 (POTENTIAL).  
 CC DOMAIN 280 287  
 CC CYTOPLASMIC (POTENTIAL).  
 CC TRANSMEM 288 308  
 CC 2 (POTENTIAL).  
 CC DOMAIN 309 339  
 CC EXTRACELLULAR (POTENTIAL).  
 CC TRANSMEM 340 360  
 CC 3 (POTENTIAL).  
 CC DOMAIN 361 381  
 CC CYTOPLASMIC (POTENTIAL).  
 CC TRANSMEM 382 402  
 CC 4 (POTENTIAL).  
 CC DOMAIN 403 421  
 CC EXTRACELLULAR (POTENTIAL).  
 CC TRANSMEM 422 442  
 CC 5 (POTENTIAL).  
 CC DOMAIN 443 469  
 CC CYTOPLASMIC (POTENTIAL).  
 CC TRANSMEM 470 490  
 CC 6 (POTENTIAL).  
 CC DOMAIN 491 532  
 CC EXTRACELLULAR (POTENTIAL).  
 CC TRANSMEM 533 553  
 CC 7 (POTENTIAL).  
 CC DOMAIN 554 1036  
 CC CYTOPLASMIC (POTENTIAL).  
 CC TRANSMEM 555 206  
 CC 8 (POTENTIAL).  
 CC DOMAIN 816 819  
 CC POLY-SER.  
 CC TRANSMEM 819 819  
 CC 9 (POTENTIAL).  
 CC CARBOHYD 95 55  
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC TRANSMEM 95 95  
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC CARBOHYD 184 184  
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC TRANSMEM 195 195  
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC CARBOHYD 213 213  
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC TRANSMEM 213 336  
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC CARBOHYD 336 336  
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).



FT CARBOHYD 419 419 N-LINKED (GLCNAC... ) (POTENTIAL).  
SQ SEQUENCE 1036 AA; 116552 MW; 7797FC71A539A87A CRC64;  
  
Query Match 78.0%; Score 32; DB 1; Length 1036;  
Best Local Similarity 71.4%; Pred. No. 85;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
Qy 1 TKHGPRK 7  
Db 659 SSHGPRK 665  
  
RESULT 3  
US20\_HUMAN  
ID US20\_HUMAN STANDARD; PRT; 1701 AA.  
AC 075643; 094884;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE US small nuclear ribonucleoprotein 200 kDa helicase (US snRNP-specific  
DE 200 kDa protein) (US-200KD) (Fragment).  
GN KIAA0788;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 237-255; 765-778; 860-882;  
RP 891-903 AND 1282-1293.  
RC TISSUE=Fetal brain;  
PX MEDLINE=96324408; PubMed=8670905;  
RA Lauber J., Fabrizio P., Feigekamp S., Lane W.S., Hartmann E.,  
RA Luehrmann R.;  
RT "The HeLa 200 kDa US snRNP-specific protein and its homologue in  
RT Saccharomyces cerevisiae are members of the DEXH-box protein family of  
RT putative RNA helicases."  
RL EMBL J. 15:4001-4015(1996).  
RN [2]  
RP SEQUENCE OF 378-1701 FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=99087487; PubMed=9872452;  
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.,  
RA Tanaka A., Kotani H., Nomura N., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human genes. XI.  
RT The complete sequences of 100 new cDNA clones from brain which code  
RT for large proteins in vitro."  
RL DNA Res. 5:277-286(1998).  
CC -!- FUNCTION: PUTATIVE RNA HELICASE INVOLVED IN THE SECOND STEP OF RNA  
CC SPLICING. MAY PROMOTE ONE OR MORE CONFORMATIONAL CHANGES IN THE  
CC DYNAMIC NETWORK OF RNA-RNA INTERACTIONS IN THE SPliceosome.  
CC -!- SUBUNIT: US snRNP CONTAINS NINE SPECIFIC PROTEINS WITH MOLECULAR  
CC WEIGHTS OF 40, 52, 100, 102, 110, 116, 200 AND 220 KDA.  
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).  
CC -!- DOMAIN: COMPOSED OF TWO SIMILAR DOMAINS.  
CC -!- SIMILARITY: BELONGS TO THE HELICASE FAMILY. SKI2 SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AB018331; BAA34508.1; -;  
DR EMBL; Z70200; CAA94089.1; -;  
DR GK; 075643; -;  
DR MW; 601664; -;  
DR GO; GO:0005681; C:spliceosome complex; IDA.  
DR GO; GO:0004004; F:ATP dependent RNA helicase activity; NAS.  
DR InterPro; IPR003593; AAA ATPase.  
DR InterPro; IPR001410; DEAD.  
DR InterPro; IPR001650; Helicase\_C.

DR InterPro; IPR004179; Sec63.  
DR Pfam; PF00270; DEAD\_2; 2.  
DR Pfam; PF02771; Helicase\_C; 2.  
DR Pfam; PF02889; Sec63; 2.  
DR SMART; SMO0382; AAA; 2.  
DR SMART; SMO0487; DEXDC; 2.  
DR SMART; SMO0490; HELIC; 2.  
DR SMART; SMO0611; SEC63; 2.  
KW Helicase; mRNA processing; mRNA splicing; Spliceosome;  
KW Nuclear protein; ATP-binding; Repeat.  
FT NON\_TER 1  
FT DOMAIN 68 469  
FT DOMAIN 915 1304  
FT NP\_BIND 68 75  
FT NP\_BIND 915 922  
FT SITE 180 183  
FT SITE 1019 1022  
FT CONFLICT 936 936 N -> S (IN REF. 2).  
FT CONFLICT 948 951 RLWQ -> BALA (IN REF. 2).  
FT CONFLICT 1112 1112 F -> Y (IN REF. 2).  
FT CONFLICT 1232 1232 L -> Q (IN REF. 2).  
FT CONFLICT 1521 1521 E -> K (IN REF. 2).  
FT CONFLICT 1526 1536 RRUPPPFSGLF -> KQLPFTTSEHI (IN REF. 2).  
FT CONFLICT 1566 1669 GRHN -> AHNY (IN REF. 2).  
SQ SEQUENCE 1701 AA; 194478 MW; E072648DCB0FDB45 CRC64;  
  
Query Match 78.0%; Score 32; DB 1; Length 1701;  
Best Local Similarity 71.4%; Pred. No. 1.4e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
Qy 1 TKHGPRK 7  
Db 1116 TKHGPRK 1122  
  
RESULT 4  
EFNA1\_HUMAN  
ID EFNA1\_HUMAN STANDARD; PRT; 205 AA.  
AC P20327;  
DT 01-SEP-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Ephrin-A1 precursor (EPH-related receptor tyrosine kinase ligand 1)  
DE (EPH-A1) (immediate early response protein B61) (Tumor necrosis  
DE factor, alpha-induced protein 4).  
GN EFNA1 OR EPGL1 OR LERK1 OR TNFAIP4.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91042512; PubMed=2233719;  
RA Holzman L.B., Marks R.M., Dixit V.M.;  
RT "A novel immediate-early response gene of endothelium is induced by  
RT cytokines and encodes a secreted protein."  
RL Mol. Cell. Biol. 10:5830-5838(1990).  
RN [2]  
RP GPI-ANCHOR.  
RX MEDLINE=95140419; PubMed=7838529;  
RA Kozlosky C.J., Maraskovsky E., McGrew J.T., Vandenbos T.,  
RA Teepe M., Lyman S.D., Srinivasan S., Fletcher F.A., Gayle R.B. III,  
RA Cerretti D.P., Beckmann M.P.;  
RT "Ligands for the receptor tyrosine kinases hek and elk: isolation of  
RT cDNAs encoding a family of proteins."  
RT Oncogene 10:239-306(1995).  
CC -!- SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASES EPHA2, EPHA4,  
CC EPHA5, EPHA6 AND EPHA7. ALSO BINDS WITH LOW AFFINITY TO EPHAL.  
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
CC -!- INDUCTION: BY TNF-ALPHA AND INTERLEUKIN-1 BETA.  
CC -!- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; M57730; AAA58388.1; -;  
 CC PIR; A36377; A36377.  
 CC Genew; HGNC:3221; EFNA1.  
 CC MIM; 191164; -;  
 CC GO; GO:0005887; C: integral to plasma membrane; TAS.  
 CC GO; GO:0005102; F: receptor binding activity; TAS.  
 CC GO; GO:0005108; F: transmembrane ephrin; TAS.  
 CC GO; GO:0007267; P: cell-cell signaling; TAS.  
 CC InterPro; IPR001799; Ephrin.  
 CC Pfam; PF00812; Ephrin; 1.  
 CC PRINTS; PR01347; Ephrin.  
 CC ProDom; PD002533; Ephrin; 1.  
 CC PROSITE; PS01299; Ephrin; 1.  
 CC Glycoprotein; GPI-anchor; Signal; Polymorphism.  
 CC SIGNAL 1 17  
 CC FT CHAIN 18 205  
 CC FT CARBOHYD 26 26  
 CC FT VARIANT 159 159  
 CC FT V -> D (IN DBSNP:4745).  
 CC FT /FTID=VAR 014791.  
 CC SQ SEQUENCE 205 AA; 23771 MW; 4FEFC6BF4C1251A9 CRC64;

Query Match 75.6%; Score 31; DB 1; Length 205;  
 Best Local Similarity 83.3%; Pred. No. 26;  
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 KHGPRK 7  
 |||||  
 Db 98 KHGPEK 103

RESULT 5

EFNA1\_MOUSE STANDARD; PRT; 205 AA.  
 AC P52793; P97331;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Ephrin-A1 precursor (EPH-related receptor tyrosine kinase ligand 1)  
 DE (LEK-1) (Immediate early response protein B61).  
 GN EFNA1 OR EPLI OR LERKI OR EPLI.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ICR;  
 RX MEDLINE=95405853; PubMed=7675446;  
 RA Takahashi H., Ikeda T.;  
 RT "Molecular cloning and expression of rat and mouse B61 gene:  
 RT implications on organogenesis.";  
 RL Oncogene 11:879-883(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BALB/c.  
 RA Morris J.C., Chiarletta A., Morris G.E., Giannotti J., Caruso A.,  
 RA Hammett D.J., Finnerty H., Turner K., Wood C.R.;  
 RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=97060319; PubMed=8903354;  
 RA Flenniken A.M., Gale N.W., Yancopoulos G.D., Wilkinson D.G.;  
 RT "Distinct and overlapping expression patterns of ligands for  
 RT Eph-related receptor tyrosine kinases during mouse embryogenesis.";  
 RL Dev. Biol. 179:382-401(1996).  
 CC -1- SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASES EPHA2, EPHA4,

CC EPHAS, EPHA6 AND EPHA7. ALSO BINDS WITH LOW AFFINITY TO EPHA1.  
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (By  
 CC similarity).  
 CC -1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; D38146; BAA07344.1; -;  
 CC EMBL; U26188; AAA67563.1; -;  
 CC EMBL; U90662; AAB50237.1; -;  
 CC MGD; MGI:103236; Efnal.  
 CC InterPro; IPR001799; Ephrin.  
 CC Pfam; PF00812; Ephrin; 1.  
 CC PRINTS; PR01347; Ephrin.  
 CC ProDom; PD002533; Ephrin; 1.  
 CC PROSITE; PS01299; Ephrin; 1.  
 CC Glycoprotein; GPI-anchor; Signal.  
 CC SIGNAL 1 17  
 CC FT CHAIN 18 205  
 CC FT CARBOHYD 26 26  
 CC FT CONFLICT 74 74  
 CC FT CONFLICT 79 79  
 CC FT CONFLICT 81 81  
 CC FT CONFLICT 91 91  
 CC FT CONFLICT 94 94  
 CC FT CONFLICT 112 112  
 CC FT CONFLICT 115 115  
 CC FT CONFLICT 138 138  
 CC FT CONFLICT 154 154  
 CC FT CONFLICT 156 156  
 CC FT CONFLICT 159 159  
 CC FT CONFLICT 181 181  
 CC FT CONFLICT 204 204  
 CC SQ SEQUENCE 205 AA; 23802 MW; 5A8F3A6E2091E868 CRC64;

Query Match 75.6%; Score 31; DB 1; Length 205;  
 Best Local Similarity 83.3%; Pred. No. 26;  
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 KHGPRK 7  
 |||||  
 Db 98 KHGPEK 103

RESULT 6

EFNA1\_RAT STANDARD; PRT; 205 AA.  
 AC P97553;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Ephrin-A1 precursor (EPH-related receptor tyrosine kinase ligand 1)  
 DE (LEK-1) (Immediate early response protein B61).  
 GN EFNA1 OR EPLI OR LERKI.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OC NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Wistar; TISSUE=Brain;  
 RX MEDLINE=95405853; PubMed=7675446;  
 RA Takahashi H., Ikeda T.;  
 RT "Molecular cloning and expression of rat and mouse B61 gene:  
 RT implications on organogenesis.";  
 RL Oncogene 11:879-883(1995).  
 CC -1- SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASES EPHA2, EPHA4,

CC EPHA5, EPHA6 AND EPHA7. ALSO BINDS WITH LOW AFFINITY TO EPHA1 (BY  
 CC SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (By  
 CC similarity).  
 CC -!- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; D38056; BAA07242.1; -;  
 CC InterPro; IPR001799; Ephrin.  
 CC Pfam; PF00812; Ephrin; 1.  
 CC PRINTS; PR01347; EPHRIN.  
 CC PRODOM; PD002533; Ephrin; 1.  
 CC PROSITE; PS01299; EPHRIN; 1.  
 CC Glycoprotein; GPI-anchor; Signal.  
 CC SIGNAL 1 17 POTENTIAL.  
 CC CHAIN 18 205 EPHRIN-A1.  
 CC FT CARBOHYD 26 26 N-LINKED (GLCNAC...) (POTENTIAL).  
 CC SEQUENCE 205 AA; 23718 MW; C86DAB3DB56A6EAD CRC64;

Query Match 75.6%; Score 31; DB 1; Length 205;  
 Best Local Similarity 83.3%; Pred. No. 26;  
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 KHGPRK 7  
 DB 98 KHGFEK 103

RESULT 7  
 YG24 YEAST STANDARD; PRT; 245 AA.  
 AC P53237;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Hypothetical 28.6 kDa protein in MUP1-SPR3 intergenic region.  
 GN YK057C.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Entian K.D., Rose M., Koetter P., Roehmer A., Sehrsam I.,  
 RA Hempel S.;  
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; Z72842; CAA37058.1; -;  
 CC EMBL; Z72843; CAA37060.1; -;  
 CC PIR; S64351; S64351.  
 CC SGD; S0003289; LST7.  
 CC GO; GO:0030120; C.vesicle coat; IDA.  
 CC GO; GO:0008565; P.protein transporter activity; IDA.  
 CC GO; GO:0006893; P.googit to plasma membrane transport; IDA.  
 CC GO; GO:0006886; P.intracellular protein transport; IDA.  
 CC GO; GO:0016192; P.vesicle-mediated transport; IDA.  
 CC Hypothetical protein.  
 CC SEQUENCE 245 AA; 28560 MW; 0AFB1028E4AAC5F8 CRC64;

Query Match 75.6%; Score 31; DB 1; Length 245;  
 Best Local Similarity 100.0%; Pred. No. 31;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 KHGPR 6  
 DB 17 KHGPR 21

RESULT 8  
 MTB6\_BACSP STANDARD; PRT; 315 AA.  
 AC P43420;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Modification methylase Bsp6I (EC 2.1.1.73) (Cytosine-specific  
 DE methyltransferase Bsp6I) (M.Bsp6I).  
 DE Bsp6IM.  
 GN Bsp6IM.  
 OS Bacillus sp. (strain RFL6).  
 OG Plasmid pXh13.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

OC NCBI\_TaxID=1409;  
 RN [1]\_  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95331576; PubMed=7607501;  
 RA Lubys A., Jamlatitis A.;  
 RT "Cloning and analysis of the plasmid-borne genes encoding the Bsp6I  
 RT restriction and modification enzymes.";  
 RL Gene 157:25-29 (1995).

CC -!- FUNCTION: THIS METHYLASE RECOGNIZES THE DOUBLE-STRANDED SEQUENCE  
 CC GNGCG. CAUSES SPECIFIC METHYLATION ON C-? ON BOTH STRANDS, AND  
 CC PROTECTS THE DNA FROM CLEAVAGE BY THE Bsp6I ENDONUCLEASE.  
 CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA cytosine = S-  
 CC adenosyl-L-homocysteine + DNA 5-methylcytosine  
 CC -!- SIMILARITY: BELONGS TO THE C5-METHYLTRANSFERASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X81638; CAA57293.1; -;  
 CC PIR; I40138; I40138.  
 CC HSSP; P05102; 6MHT.  
 CC REBASE; 3315; M.Bsp6I.  
 CC InterPro; IPR001525; C5 DNA meth.

CC Pfam; PF00145; DNA methylase; 1.  
 CC PRINTS; PR00105; C5METHYTRFRASE.  
 CC TIGRFAMs; TIGR00675; Gcm; 1.  
 CC PROSITE; PS00094; C5 MTASE 1; 1.  
 CC PROSITE; PS00095; C5 MTASE 2; 1.  
 CC Transferrase; Methyltransferase; Restriction system; Plasmid.  
 CC ACT\_SITE 73 BY SIMILARITY.

CC SEQUENCE 315 AA; 36315 MW; ED2DD748AD33E44F CRC64;  
 Query Match 75.6%; Score 31; DB 1; Length 315;  
 Best Local Similarity 85.7%; Pred. No. 40;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TKHGPRK 7  
 DB 258 TKHGIRK 264

RESULT 9  
 PO21\_POPJA STANDARD; PRT; 445 AA.  
 ID PO21\_POPJA  
 AC Q03273;

DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Retrovirus-related POL polyprotein from type I retrotransposable  
 DE element R2 [contains: Reverse transcriptase (EC 2.7.7.49);  
 DE Endonuclease] (Fragment).  
 OS Popillia japonica (Japanese beetle).  
 CC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 CC Neoptera; Endopterygota; Coleoptera; Polyphaga; Scarabaeiformia;  
 CC Scarabaeidae; Rutelinae; Popillia.  
 OX NCBI\_TaxID=7064;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93196484; PubMed=8383793;  
 RA Burke W.D., Eickbush D.G., Xiong Y., Jakubczak J.L., Eickbush T.H.;  
 RT "Sequence relationship of retrotransposable elements R1 and R2 within  
 RT and between divergent insect species.";  
 RL Mol. Biol. Evol. 10:163-185(1993).  
 CC -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate  
 CC + [DNA] (N).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L00946; AAA29784.1; -;  
 KW PIR; C47757;  
 RN Transferase; RNA-directed DNA polymerase; Transposable element;  
 KW Hydrolase; Nuclease; Endonuclease.  
 FT NON\_TER 1  
 FT DOMAIN <1 114 REVERSE TRANSCRIPTASE.  
 FT DOMAIN 115 445 NUCLEIC ACID-BINDING ENDONUCLEASE.  
 SQ SEQUENCE 445 AA; 50127 MW; 2FD15CB26E92A17D CRC64;  
 Query Match 75.6%; Score 31; DB 1; Length 445;  
 Best Local Similarity 100.0%; Pred. No. 57;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 KHGR 6  
 DB 378 KHGR 382  
 RESULT 10  
 LEC\_PAPPC  
 ID LEC\_PAPPC STANDARD; PRT; 447 AA.  
 AC P83304;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Mannose/glucose-specific lectin.  
 OS Parkia platycephala.  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 CC eurosids I; Fabales; Papaceae; Mimosoideae; Parkieae; Parkia.  
 OX NCBI\_TaxID=185447;  
 RN [1]  
 RP SEQUENCE, SUBUNIT, MASS SPECTROMETRY, AND VARIANTS VAL-70; ARG-227 AND  
 RP ASN-296.  
 RC TISSUE=Seed;  
 RX MEDLINE=21393945; PubMed=11502201;  
 RA Mann K., Farias C.M.S.A., Del Sol F.G., Santos C.F., Grangeiro T.B.,  
 RA Nagano C.S., Cavada B.S., Calvete J.J.;  
 RT "The amino-acid sequence of the glucose/mannose-specific lectin  
 RT isolated from Parkia platycephala seeds reveals three tandemly  
 RT arranged jacalin-related domains";  
 RL Eur. J. Biochem. 289:4414-4422(2001).  
 RP [2]  
 RP FUNCTION.

RA Ramos M.V., Cavada B.S., Bomfim L.R., Debray H., Mazard A.-M.,  
 RA Calvete J.J., Grangeiro T.B., Rouge P.;  
 RT "Interaction of the seed lectin from Parkia platycephala (Mimosoideae)  
 RT with carbohydrates and complex glycans.";  
 RL Protein Pept. Lett. 6:215-222(1999).  
 CC -!- FUNCTION: Mannose/glucose specific lectin. Shows agglutinating  
 CC activity against rabbit erythrocytes.  
 CC -!- SUBUNIT: Homodimer.  
 CC -!- MASS SPECTROMETRY: MW=47946; MW ERR=6; METHOD=Electrospray.  
 CC -!- MASS SPECTROMETRY: MW=47951; MW ERR=9; METHOD=MALDI.  
 CC -!- SIMILARITY: BELONGS TO THE JACALIN LECTIN FAMILY.  
 DR InterPro: IPR001229; Jacalin\_lectin.  
 DR Pfam: PF01419; Jacalin; 3.  
 KW Lectin; Mannose-binding; Repeat.  
 FT MOD\_RES 1 1 BLOCKED.  
 FT DOMAIN 1 447 3 X APPROXIMATE TANDEM REPEATS.  
 FT REPEAT 1 149 1.  
 FT REPEAT 150 295 2.  
 FT REPEAT 296 447 3.  
 FT VARIANT 70 70 I -> V.  
 FT VARIANT 227 227 K -> R.  
 FT VARIANT 296 296 D -> N.  
 SQ SEQUENCE 447 AA; 47521 MW; 8F14ED460874BBB2 CRC64;  
 Query Match 75.6%; Score 31; DB 1; Length 447;  
 Best Local Similarity 100.0%; Pred. No. 58;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 TKHG 5  
 DB 252 TKHG 256  
 RESULT 11  
 FPI\_MYTGA  
 ID FPI\_MYTGA STANDARD; PRT; 751 AA.  
 AC Q27409;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Adhesive plaque matrix protein precursor (Foot protein 1) (MGFP1)  
 DE (MGFP-1).  
 GN FPI.  
 OS Mytilus galloprovincialis (Mediterranean mussel).  
 CC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Mytiloidea;  
 CC Mytiloidea; Mytilidae; Mytilus.  
 OX NCBI\_TaxID=29158;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Foot;  
 RX MEDLINE=94318724; PubMed=8043658;  
 RA Inoue K., Odo S.;  
 RT "The adhesive protein cDNA of Mytilus galloprovincialis encodes  
 RT decapetide repeats but no hexapeptide motif.";  
 RL Biol. Bull. 186:349-355(1994).  
 CC -!- FUNCTION: PROVIDES ADHESIVENESS TO THE MUSSEL'S FOOT. MUSSELS  
 CC PRODUCE ONE OF THE STRONGEST WATER INSOLUBLE GLUE. THE MUSSEL'S  
 CC ADHESIVE IS A BUNDLE OF THREADS, CALLED A BYSSUS, FORMED BY A  
 CC FIBROUS COLLAGENOUS CORE COATED WITH ADHESIVE PROTEINS.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- TISSUE SPECIFICITY: PRODUCED BY THE BYSSAL GLAND.  
 CC -!- DOMAIN: ALMOST EXCLUSIVELY COMPOSED OF REPEATS OF A DECAPEPTIDE.  
 CC -!- PTM: PROLINES IN THE REPEAT ARE HYDROXYLATED (SINGLE OR DI-) AND  
 CC ALSO TYROSINE (THUS PRODUCING DOFA = 3,4-DIHYDROXYPHENYLALANINE).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

CC  
DR EMBL; U26057; AAA80227.1; -.  
DR EMEL; D38074; BAA07268.1; -.  
DR PIR; I51685; I51685.  
DR DR InterPro; IPR003593; AAA ATPase.  
DR InterPro; IPR001208; MCM.  
DR Pfam; PF00493; MCM; 1.  
DR ProDom; PD001041; MCM; 1.

CC	EMBL; X00130; CAA24960.2; --		
DR	P1K; A20288; GNNYCL		
DR	HSSP; Q88571; 1TME.		
DR	InterPro; IPR004080; FMDVPlcoat.		
DR	InterPro; IPR001676; Rhv.		
DR	Pfam; PF000073; rhv; 3.		
DR	PRINTS; PR01542; FMDVPlcoat.		
KW	Polyprotein; Coat protein; Core protein; Nonstructural protein; Myristate.		
KW			
FT	CHAIN	1	216
FT	CHAIN	217	286
FT	CHAIN	287	504
FT	CHAIN	505	723
FT	NONSTRUCTURAL PROTEIN VP0A.		
FT	COAT PROTEIN VP4.		
FT	COAT PROTEIN VP2.		
FT	COAT PROTEIN VP3.		

FT CHAIN 724 932 COAT PROTEIN VP1.  
 FT CHAIN 933 >1011 CORE PROTEIN P12.  
 FT LIPID 217 217 MYRISTATE (BY SIMILARITY).  
 FT NON TER 1011 1011  
 SQ SEQUENCE 1011 AA; 111535 MW; 88845600A560601 CRC64;  
 Query Match 75.6%; Score 31; DB 1; Length 1011;  
 Best Local Similarity 100.0%; Pred.No. 1.3e+02; Indels 0; Gaps 0;  
 Matches 5; Conservative 0; Mismatches 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 TKHGP 5  
 Db 975 TKHGP 979  
 RESULT 14  
 POLG\_FMDVA STANDARD; PRT; 2332 AA.  
 ID POLG\_FMDVA STANDARD; PRT; 2332 AA.  
 AC P03308; Q65039; Q65040; Q65041; Q65042; Q65043;  
 AC Q65044; Q65045; Q65046; Q65047;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-JAN-1988 (Rel. 06, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Genome polypeptide [Contains: Nonstructural protein P20A; Coat  
 proteins VP1 TO VP4; Core proteins X, P14, P41, P19; Genome-linked  
 proteins VP1 TO VP3; Picornain 3C (EC 3.4.22.28) (Protease 3C)  
 DE (P2C); RNA-directed RNA polymerase (EC 2.7.7.48)].  
 DE Foot-and-mouth disease virus (strain A12) (Aphthovirus A) (FMDV).  
 OS Foot-and-mouth disease virus (strain A12) (Aphthovirus A) (FMDV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Aphthovirus.  
 OX NCBI\_TaxID=12114;  
 OX NCBI\_TaxID=12114;  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=85211015; PubMed=2987518;  
 RA Robertson B.H., Grubman M.J., Weddell G.N., Moore D.M., Welsh J.D.,  
 RA Fischer T., Dowbenko D.J., Yansura D.G., Small B., Kleid D.G.;  
 RT "Nucleotide and amino acid sequence coding for polypeptides of  
 RT foot-and-mouth disease virus type A12.";  
 RL J. Virol. 54:651-660(1985).  
 [2]  
 RP SEQUENCE OF 1863-2332 FROM N.A.  
 RX MEDLINE=83225613; PubMed=6305004;  
 RA Robertson B.H., Morgan D.O., Moore D.M., Grubman M.J., Card J.,  
 RA Fischer T., Weddell G.N., Dowbenko D.J., Yansura D.G.;  
 RT "Identification of amino acid and nucleotide sequence of the  
 RT foot-and-mouth disease virus RNA polymerase.";  
 RL Virology 126:614-623(1983).  
 [3]  
 RP SEQUENCE OF 715-955 FROM N.A.  
 RX MEDLINE=82061853; PubMed=6272395;  
 RA Kleid D.G., Yansura D.G., Small B., Dowbenko D.J., Moore D.M.,  
 RA Grubman M.J., McKercher P.D., Morgan D.O., Robertson B.H.,  
 RA Bachrach H.L.;  
 RT "Cloned viral protein vaccine for foot-and-mouth disease: responses  
 RT in cattle and swine.";  
 RL Science 214:1125-1129(1981).  
 CC -!- CATALYTIC ACTIVITY: Selective cleavage of Gln-[Gly bond in the  
 CC poliovirus polypeptide. In other picornavirus reactions Glu may be  
 CC substituted for Gln, and Ser or Thr for Gly.  
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate +  
 CC [RNA] (N)  
 CC -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,  
 CC VP3, AND VP4.  
 CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,  
 CC VP3, AND VP4.  
 CC -!- PTM: SPECIFIC ENZYMOLOGICAL CLEAVAGES IN VIVO YIELD MATURE PROTEINS.  
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
 CC -----  
 CC EMBL; M10975; AAA42593.1; -;  
 DR EMBL; J02187; AAA42670.1; -;  
 DR MEROPS; C03.008; -;  
 DR InterPro; IPR004004; Calici\_pol\_hel.  
 DR InterPro; IPR004080; FMDVPlCoat.  
 DR InterPro; IPR001676; RVV.  
 DR InterPro; IPR000805; RNA\_helicase.  
 DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro; IPR001205; RNA\_pol\_F3D.  
 DR InterPro; IPR007094; RNA\_pol\_PSVir.  
 DR Pfam; PF00073; rhv; 3.  
 DR Pfam; PF00680; RNA\_dep\_RNA\_pol; 1.  
 DR Pfam; PF00910; RNA\_helicase; 1.  
 DR PRINTS; PR00918; CALICIVIRUSN.  
 DR PRINTS; PR01542; FMDVPlCoAT.  
 KW Polypeptide; Coat protein; Core protein; RNA-directed RNA polymerase;  
 KW Transferase; Hydrolase; Thiol protease; Nonstructural protein;  
 KW Myristate.  
 FT CHAIN 1 200 NONSTRUCTURAL PROTEIN P20A.  
 FT CHAIN 201 285 COAT PROTEIN VP4.  
 FT CHAIN 286 503 COAT PROTEIN VP2.  
 FT CHAIN 504 723 COAT PROTEIN VP3.  
 FT CHAIN 724 937 COAT PROTEIN VP1.  
 FT CHAIN 938 953 CORE PROTEIN X.  
 FT CHAIN 954 1107 CORE PROTEIN P14.  
 FT CHAIN 1108 1425 CORE PROTEIN P41.  
 FT CHAIN 1426 1578 CORE PROTEIN P19.  
 FT CHAIN 1579 1601 GENOME-LINKED PROTEIN VP1.  
 FT CHAIN 1602 1625 GENOME-LINKED PROTEIN VP2.  
 FT CHAIN 1626 1649 GENOME-LINKED PROTEIN VP3.  
 FT CHAIN 1650 1862 PROTEASE.  
 FT CHAIN 1863 2332 RNA-DIRECTED RNA POLYMERASE.  
 FT LIPID 201 201 MYRISTATE.  
 SQ SEQUENCE 2332 AA; 259408 MW; EE77DA739CEDEC6A CRC64;  
 Query Match 75.6%; Score 31; DB 1; Length 2332;  
 Best Local Similarity 100.0%; Pred.No. 3.1e+02; Indels 0; Gaps 0;  
 Matches 5; Conservative 0; Mismatches 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 TKHGP 5  
 Db 980 TKHGP 984  
 RESULT 15  
 POLG\_FMDVO STANDARD; PRT; 2332 AA.  
 ID POLG\_FMDVO STANDARD; PRT; 2332 AA.  
 AC P03305;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Genome polypeptide [Contains: Nonstructural protein P20A; Coat  
 protein VP1; Coat protein VP2; Coat protein VP3; Coat protein VP4;  
 DE Core protein P12; Core protein P14; Core protein P19; Genome-linked  
 DE protein VP1; Protease (EC 3.4.22.-); RNA-directed RNA polymerase  
 DE (EC 2.7.7.48)].  
 OS Foot-and-mouth disease virus (strain O1) (Aphthovirus O) (FMDV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
 OC Aphthovirus.  
 OX NCBI\_TaxID=73482;  
 OX NCBI\_TaxID=73482;  
 RN SEQUENCE FROM N.A.  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O1K;  
 RX MEDLINE=84297249; PubMed=6089122;  
 RA Foras S., Strebel K., Beck E., Schaller H.;  
 RT "Nucleotide sequence and genome organization of foot-and-mouth  
 RT disease virus";  
 RL Nucleic Acids Res. 12:6587-6601(1984).  
 RN [2]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN=O1BFS;



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 16:55:29 ; Search time 9.3333 Seconds  
(without alignments)  
72.127 Million cell updates/sec

Title: US-09-901-187C-6

Perfect score: 41

Sequence: 1 TKHGPRK 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168692 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 76:.\*  
1: PIR1.\*  
2: PIR2.\*  
3: PIR3.\*  
4: PIR4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36	87.8	419	2 H83965	hypothetical prote
2	34	82.9	612	2 E82756	beta-galactosidase
3	33	80.5	369	2 T45299	hypothetical prote
4	33	80.5	551	2 T16426	hypothetical prote
5	33	80.5	555	2 T43357	potassium channel
6	33	80.5	567	2 AB2564	hypothetical prote
7	33	80.5	597	2 AB2572	hypothetical prote
8	32	78.0	311	2 H82541	conserved hypothet
9	32	78.0	436	2 F82354	hypothetical prote
10	32	78.0	1024	2 S71804	receptor-like serp
11	32	78.0	1738	2 C84507	hypothetical prote
12	32	78.0	2240	2 T37057	probable multi-dom
13	31	75.6	99	2 S43073	hypothetical prote
14	31	75.6	115	2 H72768	hypothetical prote
15	31	75.6	123	2 B69351	hypothetical prote
16	31	75.6	205	2 A36377	B61 protein precur
17	31	75.6	231	2 T16160	hypothetical prote
18	31	75.6	245	2 S64351	hypothetical prote
19	31	75.6	248	2 T40149	hypothetical prote
20	31	75.6	309	2 T00503	probable MYB fami
21	31	75.6	315	2 T40138	site-specific DNA-
22	31	75.6	320	2 C85440	myb-related protei
23	31	75.6	321	2 A10007	lipopolysaccharide
24	31	75.6	326	2 H72674	hypothetical prote
25	31	75.6	375	2 C95106	carboxynorspermid
26	31	75.6	445	2 C47757	retrovirus-related
27	31	75.6	721	2 T14229	NADH2 dehydrogenas
28	31	75.6	751	2 S68957	adhesive plaque pr
29	31	75.6	807	1 I51685	replication licens

30 31 75.6 1011 1 GNNYC1  
31 31 75.6 1048 2 T30815  
32 31 75.6 2332 1 GNNYF  
33 31 75.6 2332 1 GNNY4F  
34 31 75.6 2333 1 GNNY2F  
35 31 75.6 2336 2 S37077  
36 31 75.6 3511 2 A59295  
37 30 73.2 88 2 H86833  
38 30 73.2 103 2 G42528  
39 30 73.2 117 2 H72706  
40 30 73.2 273 2 B69883  
41 30 73.2 295 2 H83989  
42 30 73.2 324 1 D42951  
43 30 73.2 353 2 S21057  
44 30 73.2 357 1 K2ECP  
45 30 73.2 357 2 C91208

#### ALIGNMENTS

##### RESULT 1

H83965

hypothetical protein BH528 [imported] - Bacillus halodurans (strain C-125)

C:Species: Bacillus halodurans

C>Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 02-Aug-2002

C:Accession: H83965

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hiz  
Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and

A:Reference number: A83650; MUID:20512582; PMID:11058132

A:Accession: H83965

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-419 <STO>

A:Cross-references: GB:AP001515; GB:BA000004; NID:g10174886; PIDN:BA06247.1; GSPDB:GNC

A:Experimental source: strain C-125

C:Genetics:

A:Gens: BH2528

C:Superfamily: hypothetical protein c0103

Query Match 87.8%; Score 36; DB 2; Length 419;

Best Local Similarity 100.0%; Pred. No. 15;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 KHGPRK 7

DB 69 KHGPRK 74

##### RESULT 2

E82756

beta-galactosidase XF0840 [imported] - Xylella fastidiosa (strain 9a5c)

C:Species: Xylella fastidiosa

C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Sep-2000

C:Accession: E82756

R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequ

Nature 406, 151-157, 2000

A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A:Reference number: A82515; MUID:20365717; PMID:10910347

A>Note: for a complete list of authors see reference number A59328 below

A:Accession: E82756

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-612 <SIM>

A:Cross-references: GB:A8003923; GB:A8003849; NID:g9105736; PIDN:AAF83650.1; GSPDB:GN00

A:Experimental source: strain 9a5c

R:Simpson, A.J.G.; Reinach, P.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.;

Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer,

as-Neto, E.; Docena, C.; El-Dorry, H.; Pacinani, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Froh

J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Lalig



Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.  
 A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.V.;  
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.  
 Rodrigues, V.; Rosa, A.J.; de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawaak  
 A;Authors: da Silva, A.C.R.; da Silva, P.R.; da Silva, A.W.; Silva Jr., W.A.; da Silva  
 M.; Tshukko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
 A;Reference number: A59328  
 A;Contents: annotation  
 C;Genetics:  
 A;Gene: XP0840  
 C;Superfamily: beta-galactosidase bga

Query Match 82.9%; Score 34; DB 2; Length 612;  
 Best Local Similarity 85.7%; Pred. No. 55;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TRHGPRK 7  
 :|||||  
 Db 413 TLHGPRK 419

## RESULT 3

hypothetical protein DKFPZ434J0310.1 - human  
 C;Species: Homo sapiens (man)  
 C;Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 04-Feb-2000  
 R;Duesterhoeft, A.; Lauber, J.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
 submitted to the Protein Sequence Database, January 2000  
 A;Reference number: Z23035  
 A;Accession: T46299  
 A;Status: preliminary  
 A;Molecule type: mRNA  
 A;Residues: 1-369 <AAA>  
 A;Cross-references: EMBL:AL137745  
 A;Experimental source: adult testis; clone DKFPZ434J0310  
 C;Genetics:  
 A;Note: DKFPZ434J0310.1

Query Match 80.5%; Score 33; DB 2; Length 369;  
 Best Local Similarity 71.4%; Pred. No. 54;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 TRHGPRK 7  
 :|||||  
 Db 224 SEHGPRK 230

## RESULT 4

hypothetical protein F52E4.4 - Caenorhabditis elegans  
 C;Species: Caenorhabditis elegans  
 C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 18-Feb-2000  
 C;Accession: T16426  
 R;Martin, J.  
 submitted to the EMBL Data Library, April 1996  
 A;Description: The sequence of C. elegans cosmid F52E4.  
 A;Reference number: Z18512  
 A;Accession: T16426  
 A;Status: preliminary; translated from GB/EMBL/DDBJ  
 A;Molecule type: DNA  
 A;Residues: 1-551 <MAR>  
 A;Cross-references: EMBL:U56964; NID:G1293826; PID:G2078439; PIDN:AA54033.1; GSPDB:GN00  
 A;Experimental source: strain Bristol N2; clone F52E4  
 C;Genetics:  
 A;Gene: CESP:F52E4.4  
 A;Map position: X  
 A;Introns: 14/3; 65/3; 95/2; 118/1; 165/3; 267/1; 297/2; 334/3; 385/2; 477/1; 513/2

Query Match 80.5%; Score 33; DB 2; Length 551;  
 Best Local Similarity 83.3%; Pred. No. 79;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 KHGPRK 7  
 :|||||  
 Db 513 KHGPRR 518

## RESULT 5

hypothetical protein alr8509 [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC7120d  
 C;Species: Nostoc sp. PCC 7120  
 A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
 C;Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002  
 C;Accession: AB2564  
 R;Kakazaki, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguch  
 DNA Res. 8, 205-213, 2001  
 A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ar  
 A;Reference number: AB1807; MUID:21595285; PMID:11759840  
 A;Accession: AB2572  
 R;Kakazaki, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguch  
 DNA Res. 8, 205-213, 2001  
 A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ar  
 A;Reference number: AB1807; MUID:21595285; PMID:11759840

Query Match 80.5%; Score 33; DB 2; Length 555;  
 Best Local Similarity 83.3%; Pred. No. 79;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 KHGPRK 7  
 :|||||  
 Db 517 KHGPRR 522

## RESULT 6

hypothetical protein alr8509 [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC7120d  
 C;Species: Nostoc sp. PCC 7120  
 A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
 C;Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002  
 C;Accession: AB2564  
 R;Kakazaki, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguch  
 DNA Res. 8, 205-213, 2001  
 A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ar  
 A;Reference number: AB1807; MUID:21595285; PMID:11759840  
 A;Accession: AB2564  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-567 <KUR>  
 A;Cross-references: GB:AP003604; PIDN:BA577428.1; PID:G17134872; GSPDB:GN00183  
 A;Experimental source: strain PCC 7120  
 C;Genetics:  
 A;Gene: alr8509  
 A;Genome: plasmid

Query Match 80.5%; Score 33; DB 2; Length 567;  
 Best Local Similarity 85.7%; Pred. No. 81;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TRHGPRK 7  
 :|||||  
 Db 64 TRHGPRK 70

## RESULT 7

hypothetical protein alr9001 [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC7120e  
 AB2572  
 C;Species: Nostoc sp. PCC 7120  
 A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
 C;Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002  
 C;Accession: AB2572  
 R;Kakazaki, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguch  
 DNA Res. 8, 205-213, 2001  
 A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ar  
 A;Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AB2572  
 A:Status: Preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-597 <KUR>  
 A:Cross-references: GB:AP003605; PIDN:BA877487.1; PID:G17134932; GSPDB:GN00184  
 A:Experimental source: strain PCC 7120  
 C:Genetics:  
 A:Gene: air9001  
 A:Genome: Plasmid

Query Match 80.5%; Score 33; DB 2; Length 597;  
 Best Local Similarity 85.7%; Pred. No. 85;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TKHGPRK 7  
 |||||  
 DB 64 TKHGPRK 70

RESULT 8  
 H82541  
 conserved hypothetical protein XF2573 [imported] - Xylella fastidiosa (strain 9a5c)  
 C:Species: Xylella fastidiosa  
 C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
 C:Accession: H82541  
 R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing  
 Nature 406, 151-157, 2000  
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
 A:Reference number: AB2515; MUID:20365717; PMID:10910347  
 A:Note: for a complete list of authors see reference number A59328 below  
 A:Accession: H82541  
 A:Status: Preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-311 <SIM>  
 A:Cross-references: GB:AB004064; GB:AB003849; NID:G9107775; PIDN:AAF85370.1; GSPDB:GN001  
 A:Experimental source: strain 9a5c  
 R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A  
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H  
 as-Neto, E.; Docena, C.; EL-Dorriy, H.; Facincani, A.P.; Ferreira, A.J.S.  
 submitted to GenBank June 2000  
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm  
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuranae, E.E.; Laig  
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E  
 A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;  
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A  
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak  
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir  
 M.; Tsuchiko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
 A:Reference number: A59328  
 A:Contents: annotation  
 C:Genetics:  
 A:Gene: XF2573

Query Match 78.0%; Score 32; DB 2; Length 311;  
 Best Local Similarity 83.3%; Pred. No. 72;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TKHGPR 6  
 |||||  
 DB 67 TQHGPR 72

RESULT 9  
 F82354  
 hypothetical protein VC0179 [imported] - Vibrio cholerae (strain N16961 serogroup O1)  
 C:Species: Vibrio cholerae  
 C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
 R:Heidelberger, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;  
 chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, B  
 l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
 Nature 406, 477-483, 2000  
 A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A:Reference number: AB2035; MUID:20406833; PMID:10952301  
 A:Accession: F82354  
 A:Status: Preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-436 <HEI>  
 A:Cross-references: GB:AB004108; GB:AB003852; NID:G9654578; PIDN:AAF93355.1; GSPDB:GN00  
 A:Experimental source: serogroup O1; strain N16961; biotype El Tor  
 C:Genetics:  
 A:Gene: VC0179  
 A:Map position: 1

Query Match 78.0%; Score 32; DB 2; Length 436;  
 Best Local Similarity 83.3%; Pred. No. 18+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 KHGPRK 7  
 |||||  
 DB 358 KHGPRE 363

RESULT 10  
 S71804  
 receptor-like serpentine protein smoothened - fruit fly (Drosophila melanogaster)  
 C:Species: Drosophila melanogaster  
 C:Date: 19-Mar-1998 #sequence\_revision 17-Apr-1998 #text\_change 17-Apr-1998  
 C:Accession: S71804  
 R:van den Heuvel, M.; Ingham, P.W.  
 Nature 382, 547-551, 1996  
 A:Title: smoothened encodes a receptor-like serpentine protein required for hedgehog si  
 A:Reference number: S71804; MUID:96320560; PMID:8700230  
 A:Accession: S71804  
 A:Status: Preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-1024 <VAN>  
 A:Note: intron-exon boundaries were confirmed by sequencing fragments of the genomic se

Query Match 78.0%; Score 32; DB 2; Length 1024;  
 Best Local Similarity 71.4%; Pred. No. 2,2e+02;  
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 TKHGPRK 7  
 |||||  
 DB 647 SSHGPRK 653

RESULT 11  
 CB4507  
 hypothetical protein At2g13370 [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 23-Dec-2002  
 C:Accession: CB4507  
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L  
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,  
 Nature 402, 761-768, 1999  
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
 A:Reference number: AB4420; MUID:20083487; PMID:10617197  
 A:Accession: CB4507  
 A:Status: Preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1738 <STO>  
 A:Cross-references: GB:AE002093; NID:G4733988; PIDN:AAD28668.1; GSPDB:GN00139  
 C:Genetics:  
 A:Gene: At2g13370  
 A:Map position: 2  
 C:Superfamily: chromodomain helicase CHD1; chromobox homology

Query Match 78.0%; Score 32; DB 2; Length 1738;  
 Best Local Similarity 85.7%; Pred. No. 3,7e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TKHGPRK 7  
 |||||

Db 1516 TKDGPBK 1522

# RESULT 12

Probable multi-domain beta keto-acyl synthase - Streptomyces coelicolor  
 C:Species: Streptomyces coelicolor  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 28-Jul-2000  
 C:Accession: T37057  
 R:Seeger, K.J.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
 submitted to the EMBL Data Library, August 1999  
 A:Reference number: Z21620  
 A:Accession: T37057  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-2240 <SEE>  
 A:Cross-references: EMBL:AL109747; PIDN:CAE52354.1; GSPDB:GN00070; SCORDB:SCJ21.08  
 A:Experimental source: strain A3(2)  
 C:Genetics:

C:Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase I homology  
 F;24-435/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS>

Query Match 78.0%; Score 32; DB 2; Length 2240;  
 Best Local Similarity 83.3%; Pred. No. 4.7e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 KHGPRK 7  
 :|||||  
 Db 1174 EHGPRK 1179

# RESULT 13

Hypothetical protein 7 - human herpesvirus 6  
 C:Species: human herpesvirus 6  
 C:Date: 07-Sep-1994 #sequence\_revision 10-Nov-1995 #text\_change 08-Oct-1999  
 C:Accession: S43073  
 R:Thompson, J.; Choudhury, S.; Kashanchi, F.; Doniger, J.; Berneman, Z.; Frenkel, N.; Ro  
 Oncogene 9, 1167-1175, 1994  
 A:Title: A transforming fragment within the direct repeat region of human herpesvirus ty  
 A:Reference number: S43067; MUID:94181269; PMID:8134119  
 A:Accession: S43073  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-99 <THO>  
 A:Cross-references: EMBL:X73675; NID:G469952; PIDN:CAA52030.1; PID:G469959  
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1993

Query Match 75.6%; Score 31; DB 2; Length 99;  
 Best Local Similarity 83.3%; Pred. No. 38;  
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TKHGPR 6  
 :|||||  
 Db 74 TSHGPR 79

# RESULT 14

Hypothetical protein APE0139 - Aeropyrum pernix (strain K1)  
 C:Species: Aeropyrum pernix  
 C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000  
 C:Accession: H72768  
 R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah  
 awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K  
 DNA Res. 6, 83-101, 1999  
 A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyz  
 A:Reference number: A72450; MUID:93310339; PMID:10382966  
 A:Accession: H72768  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-115 <KAW>

A:Cross-references: DDBJ:AF000058; NID:G5103388; PIDN:BAA79050.1; PID:G5103529  
 A:Experimental source: strain K1  
 C:Genetics:  
 A:Gene: APE0139  
 C:Superfamily: Aeropyrum pernix hypothetical protein APE0139

Query Match 75.6%; Score 31; DB 2; Length 115;  
 Best Local Similarity 83.3%; Pred. No. 44;  
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TKHGPR 6  
 :|||||  
 Db 53 TSHGPR 58

# RESULT 15

Hypothetical protein AF0810 - Archaeoglobus fulgidus  
 C:Species: Archaeoglobus fulgidus  
 C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 22-Oct-1999  
 C:Accession: B69351  
 R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodsc  
 ; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F  
 Glodok, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
 Nature 390, 364-370, 1997  
 A:Authors: Overbeek, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S  
 Smith, R.O.; Woese, C.R.; Venter, J.C.  
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae  
 A:Reference number: A69250; MUID:98049343; PMID:9389475  
 A:Accession: B69351  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-123 <KLE>  
 A:Cross-references: GB:AE001048; GB:AE000782; NID:G2689371; PIDN:AAB90433.1; PID:G26498

Query Match 75.6%; Score 31; DB 2; Length 123;  
 Best Local Similarity 100.0%; Pred. No. 47;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 HGPBK 7  
 :|||||  
 Db 110 HGPBK 114

Search completed: February 11, 2004, 17:11:40  
 Job time : 11.3933 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 17:09:56 ; Search time 24.6667 Seconds  
(without alignments)  
59.419 Million cell updates/sec

Title: US-09-901-187C-6

Perfect score: 41

Sequence: 1 TKHGPRK 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 801455 seqs, 209382283 residues

Total number of hits satisfying chosen parameters: 801455

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	41	100.0	7	US-09-901-187B-6	Sequence 6, Appli
2	36	87.8	12	US-10-369-493-17328	Sequence 17328, A
3	34	82.9	572	US-10-369-493-9267	Sequence 9267, Ap
4	34	82.9	612	US-10-369-493-17539	Sequence 17539, A
5	33	80.5	56	US-10-106-698-7718	Sequence 7718, Ap
6	33	80.5	250	US-10-378-029-71	Sequence 71, Appl
7	33	80.5	369	US-10-220-381-9	Sequence 9, Appl
8	33	80.5	443	US-10-094-749-3013	Sequence 3013, Ap
9	33	80.5	572	US-10-369-493-9495	Sequence 9495, Ap
10	33	80.5	897	US-09-815-242-12769	Sequence 12769, A
11	33	80.5	1113	US-09-815-242-5836	Sequence 5836, Ap
12	32	78.0	891	US-10-310-154-599	Sequence 599, Ap
13	32	78.0	891	US-09-906-453-4	Sequence 4, Appli
14	31	75.6	65	US-10-156-761-9796	Sequence 9796, Ap
15	31	75.6	66	US-10-083-357-1061	Sequence 1061, Ap

16	31	75.6	102	12	US-10-029-386-28618	Sequence 28618, A
17	31	75.6	121	12	US-10-291-172-630	Sequence 630, App
18	31	75.6	204	12	US-10-137-870-288	Sequence 288, App
19	31	75.6	204	12	US-10-140-018-288	Sequence 288, App
20	31	75.6	204	12	US-10-140-021-288	Sequence 288, App
21	31	75.6	204	12	US-10-140-274-288	Sequence 288, App
22	31	75.6	204	12	US-10-140-471-288	Sequence 288, App
23	31	75.6	204	12	US-10-140-807-288	Sequence 288, App
24	31	75.6	204	12	US-10-140-922-288	Sequence 288, App
25	31	75.6	204	12	US-10-140-924-288	Sequence 288, App
26	31	75.6	204	12	US-10-140-926-288	Sequence 288, App
27	31	75.6	204	12	US-10-141-698-288	Sequence 288, App
28	31	75.6	204	12	US-10-141-702-288	Sequence 288, App
29	31	75.6	204	12	US-10-141-704-288	Sequence 288, App
30	31	75.6	204	12	US-10-142-421-288	Sequence 288, App
31	31	75.6	204	12	US-10-142-432-288	Sequence 288, App
32	31	75.6	204	12	US-10-142-767-288	Sequence 288, App
33	31	75.6	204	12	US-10-143-033-288	Sequence 288, App
34	31	75.6	204	12	US-10-144-994-288	Sequence 288, App
35	31	75.6	204	12	US-10-145-628-288	Sequence 288, App
36	31	75.6	204	12	US-10-145-631-288	Sequence 288, App
37	31	75.6	204	12	US-10-145-633-288	Sequence 288, App
38	31	75.6	204	12	US-10-145-746-288	Sequence 288, App
39	31	75.6	204	12	US-10-145-748-288	Sequence 288, App
40	31	75.6	204	12	US-10-145-823-288	Sequence 288, App
41	31	75.6	204	12	US-10-145-826-288	Sequence 288, App
42	31	75.6	204	12	US-10-145-876-288	Sequence 288, App
43	31	75.6	204	12	US-10-145-959-288	Sequence 288, App
44	31	75.6	204	12	US-10-146-724-288	Sequence 288, App
45	31	75.6	204	12	US-10-146-724-288	Sequence 288, App

#### ALIGNMENTS

#### RESULT 1

US-09-901-187B-6

; Sequence 6, Application US/09901187B  
; Patent No. US20020151464A1  
; GENERAL INFORMATION:  
; APPLICANT: Panacea Pharmaceuticals, Inc.  
; APPLICANT: Wolozin, Benjamin  
; APPLICANT: Ostretova-Golts, Natalie  
; APPLICANT: Lebowitz, Michael S.  
; TITLE OF INVENTION: Methods for Preventing Neural Tissue Damage and for the Treatment  
; TITLE OF INVENTION: Alpha-Synuclein Diseases  
; FILE REFERENCE: PANO1/002US  
; CURRENT FILING DATE: 2001-07-09  
; PRIOR FILING DATE: 2001-07-09  
; PRIOR APPLICATION NUMBER: US 60/217,319  
; PRIOR FILING DATE: 2000-07-07  
; PRIOR APPLICATION NUMBER: US 60/279,199  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 6  
; LENGTH: 7  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-901-187B-6

Query Match 100.0%; Score 41; DB 10; Length 7;

Best Local Similarity 100.0%; Pred. No. 7e+05;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TKHGPRK 7

DB 1 TKHGPRK 7

#### RESULT 2

US-10-369-493-17328

; Sequence 17328, Application US/10369493

Publication No. US20030233675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 17328  
LENGTH: 419  
TYPE: PRT  
ORGANISM: Bacillus halodurans  
US-10-369-493-17328

Query Match 87.8%; Score 36; DB 12; Length 419;  
Best Local Similarity 100.0%; Pred. No. 56;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 KHGPRK 7  
Db 69 KHGPRK 74

RESULT 3  
US-10-369-493-9267  
Sequence 9267, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 9267  
LENGTH: 572  
TYPE: PRT  
ORGANISM: Xylella fastidiosa  
US-10-369-493-9267

Query Match 82.9%; Score 34; DB 12; Length 572;  
Best Local Similarity 85.7%; Pred. No. 1.8e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TKHGPRK 7  
Db 413 TLHGPRK 419

RESULT 4  
US-10-369-493-17539  
Sequence 17539, Application US/10369493  
Publication No. US20030233675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng

Publication No. US20030233675A1  
GENERAL INFORMATION:  
APPLICANT: Cao, Yongwei  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
APPLICANT: Goldman, Barry S.  
APPLICANT: Chen, Xianfeng  
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
FILE REFERENCE: 38-10(52052)B  
CURRENT APPLICATION NUMBER: US/10/369,493  
CURRENT FILING DATE: 2003-02-28  
PRIOR APPLICATION NUMBER: US 60/360,039  
PRIOR FILING DATE: 2002-02-21  
NUMBER OF SEQ ID NOS: 47374  
SEQ ID NO 17539  
LENGTH: 612  
TYPE: PRT  
ORGANISM: Xylella fastidiosa  
US-10-369-493-17539

Query Match 82.9%; Score 34; DB 12; Length 612;  
Best Local Similarity 85.7%; Pred. No. 1.9e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TKHGPRK 7  
Db 413 TLHGPRK 419

RESULT 5  
US-10-106-698-7718  
Sequence 7718, Application US/10106698  
Publication No. US20030109690A1  
GENERAL INFORMATION:  
APPLICANT: Ruben et al.  
TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides  
FILE REFERENCE: PA005P1  
CURRENT APPLICATION NUMBER: US/10/106,698  
CURRENT FILING DATE: 2002-03-27  
PRIOR APPLICATION NUMBER: PCT/US00/26524  
PRIOR FILING DATE: 2000-09-28  
PRIOR APPLICATION NUMBER: US 60/157,137  
PRIOR FILING DATE: 1999-09-29  
PRIOR APPLICATION NUMBER: US 60/163,280  
PRIOR FILING DATE: 1999-11-03  
NUMBER OF SEQ ID NOS: 8564  
SOFTWARE: Patent in Ver. 3.0  
SEQ ID NO 7718  
LENGTH: 56  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: MISC\_FEATURE  
LOCATION: (48)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-10-106-698-7718

Query Match 80.5%; Score 33; DB 15; Length 56;  
Best Local Similarity 83.3%; Pred. No. 31;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TKHGPRK 6  
Db 18 TKHGPK 23

RESULT 6  
US-10-378-029-71  
Sequence 71, Application US/10378029  
Publication No. US20040014087A1  
GENERAL INFORMATION:  
APPLICANT: HODGSON, David M.; LINCOLN, Stephen E.  
APPLICANT: RUSSO, Frank D.; SPIRO, Peter A.  
APPLICANT: BANVILLE, Steve C.; BRATCHER, Shawn R.  
APPLICANT: DUFOUR, Gerard E.; COHEN, Howard J.  
APPLICANT: ROSEN, Bruce; CHALUP, Michael S.  
APPLICANT: JACKSON, Jennifer L.; JONES, Annisa L.  
APPLICANT: YU, Jimmy Y.; GREENAWALT, Lila B.  
APPLICANT: PANZER, Scott R.; ROSEBERRY LINCOLN, Ann M.

APPLICANT: WRIGHT, Rachel J.; DANIELS, Susan E.  
; TITLE OF INVENTION: MOLECULES FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: PT-1022-1 CIP  
; CURRENT APPLICATION NUMBER: US/10/378,029  
; PRIOR FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 09/980,285  
; PRIOR FILING DATE: 2001-11-30  
; PRIOR APPLICATION NUMBER: PCT/US00/15404  
; PRIOR FILING DATE: 2000-05-31  
; PRIOR APPLICATION NUMBER: US 60/147,500  
; PRIOR FILING DATE: 1999-08-05  
; PRIOR APPLICATION NUMBER: US 60/147,542  
; PRIOR FILING DATE: 1998-08-05  
; PRIOR APPLICATION NUMBER: US 60/147,541  
; PRIOR FILING DATE: 1999-08-05  
; PRIOR APPLICATION NUMBER: US 60/147,824  
; PRIOR FILING DATE: 1999-08-05  
; PRIOR APPLICATION NUMBER: US 60/147,547  
; PRIOR FILING DATE: 1999-08-05  
; PRIOR APPLICATION NUMBER: US 60/147,530  
; PRIOR FILING DATE: 1999-08-05  
; PRIOR APPLICATION NUMBER: US 60/147,536  
; PRIOR FILING DATE: 1999-08-05  
; PRIOR APPLICATION NUMBER: US 60/147,520  
; PRIOR FILING DATE: 1999-08-05  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 104  
; SOFTWARE: PERL Program  
; SEQ ID NO 71  
; LENGTH: 250  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. US20040014087A1 366739.2.orf1  
US-10-378-029-71

Query Match 80.5%; Score 33; DB 12; Length 250;  
Best Local Similarity 71.4%; Pred. No. 1.3e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 TKHGPRK 7  
:|||||  
Db 105 SEHGPRK 111

RESULT 7  
US-10-220-381-9  
; Sequence 9, Application US/10220381  
; Publication No. US20030207430A1  
; GENERAL INFORMATION:  
; APPLICANT: INCYTE GENOMICS, INC.  
; APPLICANT: TANG, Y. TOM  
; APPLICANT: LU, Dyang Aina M.  
; APPLICANT: BANDMAN, Olga  
; APPLICANT: YUE, Henry  
; APPLICANT: AZIMZAI, Yalda  
; APPLICANT: LAL, Preeti  
; APPLICANT: BAUGHN, Mariah R.  
; TITLE OF INVENTION: HUMAN ENZYME MOLECULES  
; FILE REFERENCE: PF-0763 PCT  
; CURRENT APPLICATION NUMBER: US/10/220,381  
; CURRENT FILING DATE: 2001-03-01  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: PERL Program  
; SEQ ID NO 9  
; LENGTH: 369  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. US20030207430A1 5406614CD1

US-10-220-381-9

Query Match 80.5%; Score 33; DB 12; Length 369;  
Best Local Similarity 71.4%; Pred. No. 1.9e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 TKHGPRK 7  
:|||||  
Db 224 SEHGPRK 230

#### RESULT 8

US-10-094-749-3013  
; Sequence 3013, Application US/10094749  
; Publication No. US20030219741A1  
; GENERAL INFORMATION:  
; APPLICANT: ISOGAI, TAKAO  
; APPLICANT: SUGIYAMA, TOMOYASU  
; APPLICANT: OTSUKI, TETSUJI  
; APPLICANT: WAKAMATSU, AI  
; APPLICANT: SATO, HIROYUKI  
; APPLICANT: ISHII, SHIZUKO  
; APPLICANT: YAMAMOTO, JUN-ICHI  
; APPLICANT: ISONO, YUUKO  
; APPLICANT: HIO, YURI  
; APPLICANT: OTSUKA, KAORU  
; APPLICANT: NAGAI, KEIICHI  
; APPLICANT: IRIE, RYOTARO  
; APPLICANT: TAMECHIKA, ICHIRO  
; APPLICANT: SEKI, NAOHICO  
; APPLICANT: YOSHIKAWA, TSUTOMU  
; APPLICANT: OTSUKA, MOTOYUKI  
; APPLICANT: NAGAHARI, KENJI  
; APPLICANT: MASUHO, YASUHIKO  
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA  
; FILE REFERENCE: 084335/0160  
; CURRENT APPLICATION NUMBER: US/10/094,749  
; CURRENT FILING DATE: 2002-03-12  
; PRIOR APPLICATION NUMBER: 60/350,435  
; PRIOR FILING DATE: 2002-01-24  
; PRIOR APPLICATION NUMBER: JP 2001-328381  
; NUMBER OF SEQ ID NOS: 3381  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3013  
; LENGTH: 443  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-094-749-3013

Query Match 80.5%; Score 33; DB 12; Length 443;  
Best Local Similarity 83.3%; Pred. No. 2.2e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 KHGPRK 7  
:|||||  
Db 381 KHGPRK 386

#### RESULT 9

US-10-369-493-9495  
; Sequence 9495, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10/52052/B  
; CURRENT APPLICATION NUMBER: US/10/369,493

```

; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 9495
; LENGTH: 572
; TYPE: PRT
; ORGANISM: Xylella fastidiosa
US-10-369-493-9495

```

```

Query Match      80.5%; Score 33; DB 12; Length 572;
Best Local Similarity 85.7%; Pred. No. 2.8e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy      1 TKHGPRK 7
Db      413 TIRGPRK 419

```

```

RESULT 10
US-09-815-242-12769
Sequence 12769 Application US/09815242
Patent No. US20020061569A1

```

```

; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlson, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard

```

```

; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12769
; LENGTH: 897
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12769

```

```

Query Match      80.5%; Score 33; DB 9; Length 897;
Best Local Similarity 83.3%; Pred. No. 4.4e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 TKHGPR 6
Db      307 TKHGPK 312

```

```

RESULT 11
US-09-815-242-5836
Sequence 5836 Application US/09815242
Patent No. US20020061569A1

```

```

; GENERAL INFORMATION:

```

```

; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlson, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard

```

```

; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5836
; LENGTH: 1113
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5836

```

```

Query Match      80.5%; Score 33; DB 9; Length 1113;
Best Local Similarity 83.3%; Pred. No. 5.4e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 TKHGPR 6
Db      523 TKHGPK 528

```

```

RESULT 12
US-10-310-154-599
Sequence 599 Application US/10310154
Publication No. US20030233670A1

```

```

; GENERAL INFORMATION:

```

```

; APPLICANT: Edgerton, Michael D
; APPLICANT: Chomet, Paul S.
; APPLICANT: Adams, Thomas H.
; APPLICANT: Ruff, Thomas G.
; APPLICANT: Agarwal, Ameeta K.
; APPLICANT: Ahrens, Jeffrey E.
; APPLICANT: Ball, James A.
; APPLICANT: Banu, G.
; APPLICANT: Beil, Erin
; APPLICANT: Boddupalli, Raghava
; APPLICANT: Deikman, Jill
; APPLICANT: Deng, Molian
; APPLICANT: Dong, Jinzhao
; APPLICANT: Duff, Stephen M.
; APPLICANT: Galligan, Meghan M.
; APPLICANT: Hinchey, Brenda S.
; APPLICANT: Huang, Shihshien
; APPLICANT: Johnson, G. Richard
; APPLICANT: Jung, Vincent
; APPLICANT: Kretzmer, Keith A.
; APPLICANT: Laccetti, Lucille B.
; APPLICANT: Lai, Chao-Qiang
; APPLICANT: Lee, Gary
; APPLICANT: Lin, Jie-Yi

```

APPLICANT: Liu, Jingdong  
APPLICANT: Lu, Bin  
APPLICANT: Luethy, Michael M.  
APPLICANT: Lund, Adrian  
APPLICANT: Madson, Linda L.  
APPLICANT: Malloy, Kathleen A.  
APPLICANT: McKiel, Christine L.  
APPLICANT: Miller, Philip W.  
APPLICANT: Padmavathi, Manthikanti  
APPLICANT: Parnell, Laurence D.  
APPLICANT: Start, William G.  
APPLICANT: Tennessee, Pan  
APPLICANT: Vidya, K.R.  
APPLICANT: Wang, Haiyun  
APPLICANT: Xu, Zhanguo  
APPLICANT: Xu, Nanfei  
APPLICANT: Yang, Chunzhi  
APPLICANT: Zeng, Xiaoping  
APPLICANT: Zhang, Qiang  
APPLICANT: Zhao, Yajuan  
APPLICANT: Zhou, Li  
TITLE OF INVENTION: Gene Sequences and Uses Thereof in Plants  
FILE REFERENCE: 38-15(52796)B  
CURRENT APPLICATION NUMBER: US/10/310,154  
CURRENT FILING DATE: 2002-12-04  
PRIOR APPLICATION NUMBER: 60/337,358  
PRIOR FILING DATE: 2001-12-04  
NUMBER OF SEQ ID NOS: 736  
SEQ ID NO 599  
LENGTH: 891  
TYPE: PRT  
ORGANISM: Zea mays  
US-10-310-154-599

Query Match 78.0%; Score 32; DB 12; Length 891;  
Best Local Similarity 83.3%; Pred. No. 6.8e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TKHGPR 6  
DB 428 SKHGPR 433

RESULT 13  
US-09-906-453-4  
Sequence 4, Application US/09906453  
Patent No. US20020120125A1  
GENERAL INFORMATION:  
APPLICANT: Kaepler, Shawn  
APPLICANT: Springer, Nathan  
APPLICANT: Helentjaris, Timothy  
APPLICANT: Phillips, Ronald  
TITLE OF INVENTION: Polycomb genes from Maize - Mez1 and Mez2  
FILE REFERENCE: WISCONSIN100US  
CURRENT APPLICATION NUMBER: US/09/906,453  
CURRENT FILING DATE: 2001-07-16  
NUMBER OF SEQ ID NOS: 20  
SOFTWARE: PatentIn ver. 2.1  
SEQ ID NO 4  
LENGTH: 893  
TYPE: PRT  
ORGANISM: Zea mays  
US-09-906-453-4

Query Match 78.0%; Score 32; DB 10; Length 893;  
Best Local Similarity 83.3%; Pred. No. 6.8e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TKHGPR 6  
DB 430 SKHGPR 435

RESULT 14  
US-10-156-761-9796  
Sequence 9796, Application US/10156761  
Publication No. US20030119018A1  
GENERAL INFORMATION:  
APPLICANT: OMURA, SATOSHI  
APPLICANT: IKEDA, HARUO  
APPLICANT: ISHIKAWA, JUN  
APPLICANT: HORIKAWA, HIROSHI  
APPLICANT: SHIBA, TADAYOSHI  
APPLICANT: SAKAKI, YOSHIYUKI  
APPLICANT: HATTORI, MASAHIRA  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-262  
CURRENT APPLICATION NUMBER: US/10/156,761  
CURRENT FILING DATE: 2002-05-29  
PRIOR APPLICATION NUMBER: JP 2001-204089  
PRIOR FILING DATE: 2001-05-30  
PRIOR APPLICATION NUMBER: JP 2001-272697  
PRIOR FILING DATE: 2001-08-02  
NUMBER OF SEQ ID NOS: 15109  
SEQ ID NO 9796  
LENGTH: 65  
TYPE: PRT  
ORGANISM: Streptomyces avermitilis  
US-10-156-761-9796

Query Match 75.6%; Score 31; DB 15; Length 65;  
Best Local Similarity 100.0%; Pred. No. 86;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 KHGPR 6  
DB 31 KHGPR 35

RESULT 15  
US-10-083-357-1061  
Sequence 1061, Application US/10083357  
Publication No. US20030054370A1  
GENERAL INFORMATION:  
APPLICANT: Qiangdong Zeng et al.  
TITLE OF INVENTION: Systemic Discovery of New Genes  
FILE REFERENCE: 032796-090  
CURRENT APPLICATION NUMBER: US/10/083,357  
CURRENT FILING DATE: 2002-02-27  
NUMBER OF SEQ ID NOS: 1346  
SEQ ID NO 1061  
LENGTH: 66  
TYPE: PRT  
ORGANISM: Saccharomyces cerevisiae  
US-10-083-357-1061

Query Match 75.6%; Score 31; DB 15; Length 66;  
Best Local Similarity 100.0%; Pred. No. 88;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 KHGPR 6  
DB 29 KHGPR 33

Search completed: February 11, 2004, 17:54:09  
Job time: 25.6667 secs



Gencore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 16:56:09 ; Search time 10.4167 Seconds  
(without alignments)  
28.433 Million cell updates/sec

Title: US-09-901-187C-6  
Perfect score: 41  
Sequence: 1 TKHGPRK 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgm2\_6/ptodata/1/iaa/5A COMB.pep.\*  
2: /cgm2\_6/ptodata/1/iaa/5B COMB.pep.\*  
3: /cgm2\_6/ptodata/1/iaa/6A COMB.pep.\*  
4: /cgm2\_6/ptodata/1/iaa/6B COMB.pep.\*  
5: /cgm2\_6/ptodata/1/iaa/PCTUS COMB.pep.\*  
6: /cgm2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	32	78.0	1036	2	US-08-720-484A-5
2	32	78.0	1036	3	US-08-953-823A-5
3	32	78.0	1036	4	US-09-398-239-5
4	32	78.0	1036	4	US-09-560-876A-5
5	32	78.0	1065	4	US-09-560-876A-6
6	31	75.6	205	1	US-08-321-162-2
7	31	75.6	205	1	US-08-448-736-1
8	31	75.6	205	1	US-08-441-216-2
9	31	75.6	205	1	US-08-452-779-1
10	31	75.6	205	1	US-08-299-567-4
11	31	75.6	205	2	US-08-445-065-1
12	31	75.6	205	2	US-08-445-065-12
13	31	75.6	205	3	US-08-959-524-1
14	31	75.6	205	3	US-08-959-524-12
15	31	75.6	205	4	US-09-214-631-10
16	31	75.6	205	4	US-09-611-452A-1
17	31	75.6	205	4	US-09-611-452A-12
18	31	75.6	550	3	US-09-166-460-21
19	31	75.6	550	3	US-09-361-718-21
20	31	75.6	561	4	US-09-252-991A-26010
21	31	75.6	2318	3	US-09-091-219-24
22	31	75.6	2318	4	US-09-660-541-24
23	30	73.2	51	4	US-09-345-236B-141
24	30	73.2	205	4	US-09-252-991A-25573
25	30	73.2	228	4	US-09-252-991A-27364
26	30	73.2	256	4	US-09-252-991A-32307
27	30	73.2	276	4	US-09-252-991A-32703

ALIGNMENTS

RESULT 1  
US-08-720-484A-5  
; Sequence 5, Application US/08720484A  
; Patent No. 5990281  
; GENERAL INFORMATION:  
; APPLICANT: DesSavage, Frederic  
; APPLICANT: Rosenthal, Arnon  
; APPLICANT: Stone, Donna  
; TITLE OF INVENTION: Vertebrate Smoothed Proteins  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 1 DNA Way  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080

COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WinPatIn (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/720,484A  
FILING DATE: 30-Sep-1996  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Svoboda, Craig G.  
REGISTRATION NUMBER: 39,044  
REFERENCE/DOCKET NUMBER: P1050  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650/225-1489  
TELEFAX: 650/952-9881  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1036 amino acids  
TYPE: Amino Acid  
TOPOLOGY: Linear

US-08-720-484A-5  
Sequence 17272, A  
Sequence 17549, A  
Sequence 23124, A  
Sequence 24982, A  
Sequence 30909, A  
Sequence 27860, A  
Sequence 23762, A  
Sequence 21520, A  
Sequence 22004, A  
Sequence 21352, A  
Sequence 1362, Ap  
Sequence 1371, Ap  
Sequence 1372, Ap  
Sequence 21350, A  
Sequence 16750, A  
Sequence 25407, A  
Sequence 19359, A

Query Match 78.0%; Score 32; DB 2; Length 1036;  
Best Local Similarity 71.4%; Pred. No. 1.7e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 TKHGPRK 7  
Db 659 SSHGPRK 665

RESULT 2

Thu Feb 12 09:07:46 2004

```
US-08-953-823A-5
; Sequence 5, Application US/0893823A
; Patent No. 6136958
; GENERAL INFORMATION:
; APPLICANT: DeSavage, Frederic
; APPLICANT: Rosenthal, Arnon
; APPLICANT: Stone, Donna
; TITLE OF INVENTION: Vertebrate Smoothed Proteins
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/953,823A
; FILING DATE: 30-Sep-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/027070
; FILING DATE: 30-Sep-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1050R1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1036 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
;
US-08-953-823A-5
Query Match 78.0%; Score 32; DB 3; Length 1036;
Best Local Similarity 71.4%; Pred. No. 1.7e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 TKHGPRK 7
Db 659 SSHGPRK 665

RESULT 3
US-09-398-239-5
; Sequence 5, Application US/09398239
; Patent No. 6407216
; GENERAL INFORMATION:
; APPLICANT: DeSavage, Frederic
; APPLICANT: Rosenthal, Arnon
; APPLICANT: Stone, Donna
; TITLE OF INVENTION: Vertebrate Smoothed Proteins
; FILE REFERENCE: P1050D1
; CURRENT APPLICATION NUMBER: US/09/398,239
; CURRENT FILING DATE: 1999-09-20
; EARLIER APPLICATION NUMBER: US 08/720,484
; EARLIER FILING DATE: 1996-09-30
; NUMBER OF SEQ ID NOS: 5
; SEQ ID NO 5
; LENGTH: 1036
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
;
US-09-398-239-5
Query Match 78.0%; Score 32; DB 4; Length 1036;
Best Local Similarity 71.4%; Pred. No. 1.7e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 TKHGPRK 7
Db 659 SSHGPRK 665

RESULT 4
US-09-560-876A-5
; Sequence 5, Application US/09560876A
; Patent No. 6492139
; GENERAL INFORMATION:
; APPLICANT: de Sauvage, Frederic
; APPLICANT: Rosenthal, Arnon
; APPLICANT: Stone, Donna
; TITLE OF INVENTION: Vertebrate Smoothed Proteins
; FILE REFERENCE: P1050R1D1
; CURRENT APPLICATION NUMBER: US/09/560,876A
; CURRENT FILING DATE: 2002-05-22
; PRIOR APPLICATION NUMBER: US 60/027,070
; PRIOR FILING DATE: 1996-09-30
; PRIOR APPLICATION NUMBER: US 08/953,823
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 6
; SEQ ID NO 5
; LENGTH: 1036
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
;
US-09-560-876A-5
Query Match 78.0%; Score 32; DB 4; Length 1036;
Best Local Similarity 71.4%; Pred. No. 1.7e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 TKHGPRK 7
Db 659 SSHGPRK 665

RESULT 5
US-09-560-876A-6
; Sequence 6, Application US/09560876A
; Patent No. 6492139
; GENERAL INFORMATION:
; APPLICANT: de Sauvage, Frederic
; APPLICANT: Rosenthal, Arnon
; APPLICANT: Stone, Donna
; TITLE OF INVENTION: Vertebrate Smoothed Proteins
; FILE REFERENCE: P1050R1D1
; CURRENT APPLICATION NUMBER: US/09/560,876A
; CURRENT FILING DATE: 2002-05-22
; PRIOR APPLICATION NUMBER: US 60/027,070
; PRIOR FILING DATE: 1996-09-30
; PRIOR APPLICATION NUMBER: US 08/953,823
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 6
; SEQ ID NO 6
; LENGTH: 1065
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 3
; OTHER INFORMATION: unknown amino acid
;
US-09-560-876A-6
Query Match 78.0%; Score 32; DB 4; Length 1065;
Best Local Similarity 71.4%; Pred. No. 1.8e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 TKHGPRK 7
Db 659 SSHGPRK 665

RESULT 6
US-09-560-876A-7
; Sequence 7, Application US/09560876A
; Patent No. 6492139
; GENERAL INFORMATION:
; APPLICANT: de Sauvage, Frederic
; APPLICANT: Rosenthal, Arnon
; APPLICANT: Stone, Donna
; TITLE OF INVENTION: Vertebrate Smoothed Proteins
; FILE REFERENCE: P1050R1D1
; CURRENT APPLICATION NUMBER: US/09/560,876A
; CURRENT FILING DATE: 2002-05-22
; PRIOR APPLICATION NUMBER: US 60/027,070
; PRIOR FILING DATE: 1996-09-30
; PRIOR APPLICATION NUMBER: US 08/953,823
; PRIOR FILING DATE: 1997-09-30
; NUMBER OF SEQ ID NOS: 7
; SEQ ID NO 7
; LENGTH: 1065
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 3
; OTHER INFORMATION: unknown amino acid
;
US-09-560-876A-7
Query Match 78.0%; Score 32; DB 4; Length 1065;
Best Local Similarity 71.4%; Pred. No. 1.8e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 TKHGPRK 7
Db 659 SSHGPRK 665
```

Db 688 SSHGPRK 694

RESULT 6

US-08-321-162-2  
; Sequence 2, Application US/08321162  
; Patent No. 5593669  
; GENERAL INFORMATION:  
; APPLICANT: Dixit, Vishva M.  
; TITLE OF INVENTION: Cytokine-Induced Marker for  
; TITLE OF INVENTION: Inflammatory Response  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Harness, Dickey & Pierce  
; STREET: 5445 Corporate Drive  
; CITY: Troy  
; STATE: Michigan  
; COUNTRY: USA  
; ZIP: 48098  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 720 Kb storage  
; COMPUTER: IBM PC/XT/AT  
; OPERATING SYSTEM: PC-DOS 3.3  
; SOFTWARE: ASCII TEXT EDITOR  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/321,162  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/164,611  
; FILING DATE:  
; APPLICATION NUMBER: US/07/607,741  
; FILING DATE: 16 October 1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lewak, Anna M.  
; REGISTRATION NUMBER: 33,006  
; REFERENCE/DOCKET NUMBER: 2115-00590  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (313) 641-1600  
; TELEFAX: (313) 641-0270  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 205 amino acids; mature protein 187 amino  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: unknown  
; MOLECULE TYPE: peptide  
; DESCRIPTION: Peptide and protein  
; HYPOTHETICAL: yes  
; ANTI-SENSE:  
; FRAGMENT TYPE:  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapien  
; STRAIN:  
; INDIVIDUAL ISOLATE:  
; DEVELOPMENTAL STAGE:  
; HAPLOTYPE:  
; TISSUE TYPE: HUVE  
; CELL TYPE: endothelial  
; CELL LINE: HUVE-nontransformed primary cell line  
; ORGANELLE:  
; IMMEDIATE SOURCE:  
; LIBRARY:  
; CLONE:  
; POSITION IN GENOME:  
; CHROMOSOME/SEGMENT:  
; MAP POSITION:  
; UNITS:  
; FEATURE:  
; NAME/KEY: signal peptide  
; NAME/KEY: cleavage site  
; NAME/KEY: N-linked glycosylation site

NAME/KEY: hydrophobic C-terminus  
LOCATION: +1 to +18 of Figure 3A  
LOCATION: between +18 and +19 of Figure 3A  
LOCATION: +26 to +28 of Figure 3A  
LOCATION: +182 to +205 of Figure 3A  
IDENTIFICATION METHOD: established consensus  
IDENTIFICATION METHOD: sequence  
OTHER INFORMATION: predicted mature protein 187  
OTHER INFORMATION: amino acids cleaved between +18 and +19 of Figure 3A;  
OTHER INFORMATION: molecular weight mature protein (SDS PAGE) 25kDa  
PUBLICATION INFORMATION:  
AUTHORS:  
TITLE:  
JOURNAL:  
VOLUME:  
ISSUE:  
PAGES:  
DATE:  
DOCUMENT NUMBER:  
FILING DATE:  
PUBLICATION DATE:  
RELEVANT RESIDUES IN SEQ ID NO: 2:  
US-08-321-162-2  
Query Match 75.6%; Score 31; DB 1; Length 205;  
Best Local Similarity 83.3%; Pred. No. 49;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 KHGPRK 7  
DB 98 KHGPEK 103  
RESULT 7  
US-08-448-736-1  
; Sequence 1, Application US/08448736  
; Patent No. 5650504  
; GENERAL INFORMATION:  
; APPLICANT: Bartley, Timothy Dudley  
; APPLICANT: Fox, Gary Michael  
; APPLICANT: Boyle, William James  
; APPLICANT: Welcher, Andrew Avery  
; APPLICANT: Parker, Vann Phillips  
; TITLE OF INVENTION: Eck Receptor Ligands  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Amgen Inc.  
; STREET: Amgen Center  
; STREET: 1840 Dehavilland Drive  
; CITY: Thousand Oaks  
; STATE: California  
; COUNTRY: USA  
; ZIP: 91320-1789  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 in., DS, 2.0 Mb  
; COMPUTER: Apple Macintosh  
; OPERATING SYSTEM: Macintosh OS 7.0  
; SOFTWARE: Microsoft Word Version 5.1a  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/448,736  
; FILING DATE: 24-MAY-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/145,616  
; FILING DATE: 09-NOV-1993  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 205 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single-stranded  
; TOPOLOGY: unknown  
; US-08-448-736-1

Query Match 75.6%; Score 31; DB 1; Length 205;  
Best Local Similarity 83.3%; Pred. No. 49;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 KHGPRK 7  
DB 98 KHGPEK 103

## RESULT 6

US-08-441-216-2  
Sequence 2, Application US/08441216  
Patent No. 5688656  
GENERAL INFORMATION:  
APPLICANT: DIXIT, VISHVA M.  
TITLE OF INVENTION: CYTOKINE-INDUCED MARKER FOR INFLAMMATORY  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FORSTER  
STREET: 755 Page Mill Road  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304-1018  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA: US/08/441,216  
APPLICATION NUMBER: US/08/441,216  
FILING DATE: 15-MAY-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: KOSKI, ANTOINETTE F.  
REGISTRATION NUMBER: 34,202  
REFERENCE/DOCKET NUMBER: 20344-20590.10  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 813-5600  
TELEFAX: (415) 494-0792  
TELEX: 706141  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 205 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: Protein  
US-08-441-216-2

Query Match 75.6%; Score 31; DB 1; Length 205;  
Best Local Similarity 83.3%; Pred. No. 49;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 KHGPRK 7  
DB 98 KHGPEK 103

## RESULT 9

US-08-452-779-1  
Sequence 1, Application US/08452779  
Patent No. 5716934  
GENERAL INFORMATION:  
APPLICANT: Bartley, Timothy Dudley  
APPLICANT: Fox, Gary Michael  
APPLICANT: Boyle, William James  
APPLICANT: Welcher, Andrew Avery  
APPLICANT: Parker, Vann Phillips  
TITLE OF INVENTION: Eck Receptor Ligands  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Amgen Inc.

STREET: Amgen Center  
STREET: 1840 Dehavilland Drive  
CITY: Thousand Oaks  
STATE: California  
COUNTRY: USA  
ZIP: 91320-1789  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 in., DS, 2.0 Mb  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Macintosh OS 7.0  
SOFTWARE: Microsoft Word Version 5.1a  
CURRENT APPLICATION DATA: US/08/452,779  
APPLICATION NUMBER: US/08/452,779  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/145,616  
FILING DATE:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 205 amino acids  
TYPE: amino acid  
STRANDEDNESS: single-stranded  
TOPOLOGY: unknown  
US-08-452-779-1

Query Match 75.6%; Score 31; DB 1; Length 205;  
Best Local Similarity 83.3%; Pred. No. 49;  
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 KHGPRK 7  
DB 98 KHGPEK 103

## RESULT 10

US-08-299-567-4  
Sequence 4, Application US/08299567  
Patent No. 5747033  
GENERAL INFORMATION:  
APPLICANT: Davis, et al.  
TITLE OF INVENTION: METHOD OF ENHANCING THE BIOLOGICAL  
ACTIVITY OF EPH FAMILY LIGANDS  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Regeneron Pharmaceuticals, Inc.  
STREET: 777 Old Saw Mill River Road  
CITY: Tarrytown  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10591-6707  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA: US/08/299,567  
APPLICATION NUMBER: US/08/299,567  
FILING DATE: 01-SEP-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Kempler, Gail M.  
REGISTRATION NUMBER: 32,143  
REFERENCE/DOCKET NUMBER: REG 290  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 914-345-7400  
TELEFAX: 914-345-7721  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 205 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown

```

; MOLECULE TYPE: protein
US-08-299-567-4

Query Match 75.6%; Score 31; DB 1; Length 205;
Best Local Similarity 83.3%; Pred. No. 49;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 KHGPRK 7
Db 98 KHGPEK 103

RESULT 11
US-08-445-065-1
; Sequence 1, Application US/08445065
; Patent No. 5824303
; GENERAL INFORMATION:
; APPLICANT: Bartley, Timothy D.
; APPLICANT: Boyle, William J.
; APPLICANT: Parker, Vann P.
; APPLICANT: Fox, Gary M.
; APPLICANT: Welcher, Andrew A.
; APPLICANT: Magal, Ella
; APPLICANT: Lindberg, Richard A.
; TITLE OF INVENTION: ECK Receptor Ligands
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc.
; STREET: 1840 Dehavilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 91320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/445,065
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Winter, Robert B.
; REFERENCE/DOCKET NUMBER: A-213-CIP2
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 205 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-445-065-1

Query Match 75.6%; Score 31; DB 2; Length 205;
Best Local Similarity 83.3%; Pred. No. 49;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 KHGPRK 7
Db 98 KHGPEK 103

RESULT 12
US-08-445-065-12
; Sequence 12, Application US/08445065
; Patent No. 5824303
; GENERAL INFORMATION:
; APPLICANT: Bartley, Timothy D.
; APPLICANT: Boyle, William J.
; APPLICANT: Parker, Vann P.
; APPLICANT: Fox, Gary M.
; APPLICANT: Welcher, Andrew A.
; TITLE OF INVENTION: ECK Receptor Ligands
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc.
; STREET: 1840 Dehavilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 91320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/445,065
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Winter, Robert B.
; REFERENCE/DOCKET NUMBER: A-213-CIP2
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 205 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-445-065-12

Query Match 75.6%; Score 31; DB 2; Length 205;
Best Local Similarity 83.3%; Pred. No. 49;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 KHGPRK 7
Db 98 KHGPEK 103

RESULT 13
US-08-959-524-1
; Sequence 1, Application US/08959524
; Patent No. 6081167
; GENERAL INFORMATION:
; APPLICANT: Bartley, Timothy D.
; APPLICANT: Boyle, William J.
; APPLICANT: Parker, Vann P.
; APPLICANT: Fox, Gary M.
; APPLICANT: Welcher, Andrew A.
; APPLICANT: Magal, Ella
; APPLICANT: Lindberg, Richard A.
; TITLE OF INVENTION: ECK Receptor Ligands
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc.
; STREET: 1840 Dehavilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 91320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/959,524
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/445,065
; FILING DATE:
; MOLECULE TYPE: protein
US-08-445-065-12

Query Match 75.6%; Score 31; DB 2; Length 205;
Best Local Similarity 83.3%; Pred. No. 49;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 KHGPRK 7
Db 98 KHGPEK 103
```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Winter, Robert B.
; REFERENCE/DOCKET NUMBER: A-213-CIP2
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 205 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-959-524-1
;
; Query Match 75.6%; Score 31; DB 3; Length 205;
; Best Local Similarity 83.3%; Pred. No. 49;
; Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
;
Qy 2 KHGPRK 7
Db 98 KHGPEK 103
;
; RESULT 14
US-08-959-524-12
; Sequence 12, Application US/08959524
; Patent No. 6087167
; GENERAL INFORMATION:
; APPLICANT: Bartley, Timothy D.
; APPLICANT: Boyle, William J.
; APPLICANT: Parker, Vann P.
; APPLICANT: Fox, Gary M.
; APPLICANT: Weicher, Andrew A.
; APPLICANT: Magal, Elia
; APPLICANT: Lindberg, Richard A.
; TITLE OF INVENTION: ECK Receptor Ligands
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc.
; STREET: 1840 DeBavilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 91320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/959,524
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/445,065
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Winter, Robert B.
; REFERENCE/DOCKET NUMBER: A-213-CIP2
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 205 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-959-524-12
;
; Query Match 75.6%; Score 31; DB 3; Length 205;
; Best Local Similarity 83.3%; Pred. No. 49;
; Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
;
Qy 2 KHGPRK 7
Db 98 KHGPEK 103
;
; RESULT 15
US-09-214-631-10
; Sequence 10, Application US/09214631
; Patent No. 6413730
; GENERAL INFORMATION:
; APPLICANT: Holland, Sacha
; APPLICANT: Mdamalu, Geraldine
; APPLICANT: Pawson, Tony
; TITLE OF INVENTION: OLIGOMERIZED RECEPTORS WHICH AFFECT PATHWAYS REGULATED
; BY TRANSMEMBRANE LIGANDS FOR ELK-RELATED RECEPTOR
; TITLE OF INVENTION: TYROSINE KINASES
; FILE REFERENCE: 11757.23USWO
; CURRENT APPLICATION NUMBER: US/09/214,631
; CURRENT FILING DATE: 1999-03-12
; EARLIER APPLICATION NUMBER: PCT/CA97/00473
; EARLIER FILING DATE: 1997-07-04
; EARLIER APPLICATION NUMBER: 60/021,272
; EARLIER FILING DATE: 1996-07-05
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 205
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-214-631-10
;
; Query Match 75.6%; Score 31; DB 4; Length 205;
; Best Local Similarity 83.3%; Pred. No. 49;
; Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
;
Qy 2 KHGPRK 7
Db 98 KHGPEK 103
;
; Search completed: February 11, 2004, 17:13:36
; Job time : 11.4167 secs
```

Thu Feb 12 09:07:45 2004

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 11, 2004, 16:51:33 ; Search time 32.25 Seconds  
(without alignments)  
34.452 Million cell updates/sec

Title: US-09-901-187C-6

Perfect score: 41

Sequence: 1 TKHGRP 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq 19Jun03.\*

- 1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*
- 2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*
- 3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*
- 4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*
- 5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*
- 6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*
- 7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*
- 8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*
- 9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*
- 10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*
- 11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*
- 12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*
- 13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*
- 14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*
- 15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*
- 16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*
- 17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*
- 18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*
- 19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*
- 20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*
- 21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*
- 22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*
- 23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*
- 24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	100.0	7	AAE14551	Human alpha-synuclein
2	38	92.7	66	AAU57309	Propionibacterium
3	33	80.5	56	AAU57309	Human colon cancer
4	33	80.5	83	ABH15703	Human nervous syst
5	33	80.5	150	AAAG27027	Zea mays protein I
6	33	80.5	304	AAAY49421	Human TREP1h (Exo
7	33	80.5	369	AAAG7134	Amino acid sequenc
8	33	80.5	373	ABU52987	Human nucleic acid
9	33	80.5	526	ABG15782	Novel human diagno

10	33	80.5	526	22	ABG17605	Novel human diagno
11	33	80.5	717	24	ABJ18912	Pathogen specific
12	33	80.5	743	22	ABU52986	Human nucleic acid
13	33	80.5	774	23	AAAG80778	Human exonuclease
14	33	80.5	897	22	AAU37176	Staphylococcus aur
15	33	80.5	1113	22	AAU34340	Staphylococcus aur
16	33	80.5	1637	24	ABU19113	Pathogen specific
17	32	78.0	72	23	ABP09507	Human GRP protein
18	32	78.0	127	22	AAU57616	Propionibacterium
19	32	78.0	138	22	AAU83962	Human immune/haema
20	32	78.0	147	22	AAU83078	Human immune/haema
21	32	78.0	166	22	AAU20891	Human novel foetal
22	32	78.0	174	22	AAU34889	Human protein sequ
23	32	78.0	199	22	AAU62993	Propionibacterium
24	32	78.0	890	20	AAU66188	Human bladder tumo
25	32	78.0	893	23	AAO14530	Protein of Mezz (M
26	32	78.0	997	22	ABG62887	Drosophila melanog
27	32	78.0	1036	20	AAW81063	Drosophila melanog
28	32	78.0	1036	22	ABU59323	Drosophila Smoothe
29	32	78.0	1036	22	AAU38647	Polyepitide sequen
30	32	78.0	1496	24	ABU99739	Novel human diagno
31	32	78.0	1582	22	ABG07696	Human ORFX ORF117
32	32	78.0	1701	21	AAU41353	Human ORFX ORF117
33	32	78.0	1701	24	ABU99736	Polyepitide sequen
34	32	78.0	1701	24	ABU99737	Polyepitide sequen
35	32	78.0	1811	22	AAU58550	Human protein sequ
36	32	78.0	1996	22	AAU73972	Human protein SEQ
37	32	78.0	1996	22	AAU73973	Human protein SEQ
38	32	78.0	2174	22	AAU78989	Human protein SEQ
39	32	78.0	2188	22	AAU78988	Human protein SEQ
40	31	75.6	57	21	AAU33331	Eucalyptus grandis
41	31	75.6	58	21	AAU50021	Arabidopsis thalia
42	31	75.6	63	23	ABP10989	Human ORFX protein
43	31	75.6	68	22	AAU86361	Human immune/haema
44	31	75.6	90	22	AAU51927	Propionibacterium
45	31	75.6	99	22	AAU44599	Propionibacterium

ALIGNMENTS

RESULT 1

AAE14551

ID AAE14551 standard; peptide; 7 AA.

AC AAE14551;

XX 17-MAY-2002 (first entry)

DT Human alpha-synuclein aggregation inhibitor #6.

DE Alpha-synuclein; inhibitor; neurodegenerative disease; Lewy body;  
KW Parkinson's disease; Alzheimer's disease; diffuse Lewy body disease;  
KW Multiple system atrophy; Hallervorden-Spatz disease; human.  
XX

OS Homo-sapiens.

PN WO200204482-A1

XX 17-JAN-2002

XX 06-JUL-2001; 2001WO-US21379.

XX 07-JUL-2000; 2000US-217319P.

XX 28-MAR-2001; 2001US-279199P.

XX (PANA-) PANACEA PHARM INC.

XX Wolozin B, Ostretova-Golts N, Lebowitz MS;

XX WPI; 2002-179695/23.

XX Determination of an agent capable of inhibiting aggregation of alpha

PT synuclein useful for treating a neurodegenerative disease involves  
 PT determining aggregation of alpha synuclein in the presence of exogenous  
 PT iron or copper

XX Claim 40; Page 37; 52pp; English.

CC The invention relates to screening of inhibitors of alpha-synuclein  
 CC aggregation in the presence of exogenous iron or copper. The inhibitors  
 CC are magnesium and alpha-synuclein binding peptides, which are  
 CC useful for treating neurodegenerative disease that involves  
 CC the formation of Lewy bodies e.g. Parkinson's disease (PD), Alzheimer's  
 CC disease (AD), diffuse Lewy body disease, mixed AD-PD, multiple system  
 CC atrophy and Hallerorden-Spatz disease. The present sequence is a  
 CC peptide that binds to C-terminal portion of human alpha-synuclein and  
 CC inhibits its aggregation.

XX Sequence 7 AA;

Query Match 100.0%; Score 41; DB 23; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 9.3e+05;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TKHGPRK 7  
 DB 1 TKHGPRK 7

RESULT 2  
 AAU57309  
 ID AAU57309 standard; Protein; 66 AA.

XX AAU57309;

XX 27-FEB-2002 (first entry)

DE Propionibacterium acnes immunogenic protein #18205.

KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 KW dermatological; osteopathic; neuroprotectant.

OS Propionibacterium acnes.

XX WO200181581-A2.

XX 01-NOV-2001.

XX 20-APR-2001; 2001WO-US12865.

XX 21-APR-2000; 2000US-199047P.

XX 02-JUN-2000; 2000US-208841P.

XX 07-JUL-2000; 2000US-216747P.

XX (CORI-) CORIXA CORP.

XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;

XX L'maisonneuve J, Zhang Y, Jen S, Carter D;

XX WPI; 2001-616774/71.

XX N-PSDB; AAS59582.

XX Propionibacterium acnes polypeptides and nucleic acids useful for  
 XX vaccinating against and diagnosing infections, especially useful for  
 XX treating acne vulgaris

PS Example 1; SEQ ID No 18504; 1069pp; English.  
 XX Sequences AAU39105-AAU6017 represent Propionibacterium acnes immunogenic  
 CC polypeptides. The proteins and their associated DNA sequences are used in  
 CC the treatment, prevention and diagnosis of medical conditions caused by  
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.

CC P. acnes is also involved in infections of bone, joints and the central  
 CC nervous system, however it is particularly involved in the inflammatory  
 CC lesions associated with acne vulgaris. A method for detecting the  
 CC presence or absence of P. acnes in a patient comprises contacting a  
 CC sample with a binding agent that binds to the proteins of the invention  
 CC and determining the amount of bound protein in the sample. The  
 CC polypeptides may be used as antigens in the production of antibodies  
 CC specific for P. acnes proteins. These antibodies can be used to  
 CC downregulate expression and activity of P. acnes polypeptides and  
 CC therefore treat P. acnes infections. The antibodies may also be used as  
 CC diagnostic agents for determining P. acnes presence, for example, by  
 CC enzyme linked immunosorbent assay (ELISA).

CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 66 AA;

Query Match 92.7%; Score 38; DB 22; Length 66;  
 Best Local Similarity 85.7%; Pred. No. 2.6;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 TKHGPRK 7  
 DB 5 TRHGPRK 11

RESULT 3  
 AAG76944  
 ID AAG76944 standard; Protein; 56 AA.

XX AAG76944;

XX 03-SEP-2001 (first entry)

DE Human colon cancer antigen protein SEQ ID NO:7708.

XX Human; colon cancer; colon cancer antigen; diagnosis; detection;  
 KW colorectal carcinoma.

OS Homo sapiens.

XX WO200122920-A2.

XX 05-APR-2001.

XX 28-SEP-2000; 2000WO-US26524.

XX 29-SEP-1999; 99US-0157137.

XX 03-NOV-1999; 99US-0163280.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Barash SC, Birse CE, Rosen CA;

XX WPI; 2001-235357/24.

XX N-PSDB; AAG76944.

XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,  
 XX useful for preventing, diagnosing and/or treating colorectal cancers -  
 XX Claim 11; Page 9087-9088; 9803pp; English.

XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon  
 CC cancer-associated nucleic acid molecules (N) and proteins (P), where  
 CC the proteins are collectively known as colon cancer antigens. The colon  
 CC cancer antigens have cytostatic activity and can be used in gene  
 CC therapy and vaccine production. N and P may be used in the prevention,  
 CC diagnosis and treatment of diseases associated with inappropriate P  
 CC expression. For example, N and P may be used to treat disorders  
 CC associated with decreased expression by rectifying mutations or deletions  
 CC in a patient's genome that affect the activity of P by expressing  
 CC inactive proteins or to supplement the patients own production of P.



CC Additionally, N may be used to produce the colon cancer-associated ps,  
 CC by inserting the nucleic acids into a host cell and culturing the cell  
 CC to express the proteins. N and P can be used in the prevention, diagnosis  
 CC and treatment of colorectal carcinomas and cancers. AA37196 to AA37204  
 CC and AA37789 represent sequences used in the exemplification of the  
 CC present invention.  
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were  
 CC missing at time of publication, meaning no sequences are present for  
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.

SQ Sequence 56 AA;

Query Match 80.5%; Score 33; DB 22; Length 56;

Best Local Similarity 83.3%; Pred. No. 23;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 TKHGPR 6

Db 18 TKHGPK 23

RESULT 4

ID ABB15703 standard; Protein; 83 AA.

AC ABB15703;

DT 23-JAN-2002 (first entry)

DE Human nervous system related polypeptide SEQ ID NO 4360.

XX Human; neotropic; neuroprotective; cytostatic; dermatological; virucide;  
 KW immunosuppressive; antineoplastic; anti-HIV; antibacterial; vulnary;  
 KW antiparkinsonian; anticlinal; antianaemic; antichronic; cancer;  
 KW antirheumatic; hepatotropic; cerebroprotective; antineoplastic;  
 KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;  
 KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; nephrotropic; gene therapy; vaccine.

XX Homo sapiens.

XX WO200159063-A2.

XX 16-AUG-2001.

XX 17-JAN-2001; 2001WO-US01334.

XX 31-JAN-2000; 2000US-0179065.

XX 04-FEB-2000; 2000US-0180628.

XX 24-FEB-2000; 2000US-0184664.

XX 02-MAR-2000; 2000US-0186350.

XX 16-MAR-2000; 2000US-0189874.

XX 17-MAR-2000; 2000US-0190076.

XX 18-APR-2000; 2000US-0198123.

XX 19-MAY-2000; 2000US-0205515.

XX 07-JUN-2000; 2000US-0209467.

XX 28-JUN-2000; 2000US-0214886.

XX 30-JUN-2000; 2000US-0215135.

XX 07-JUL-2000; 2000US-0216647.

XX 07-JUL-2000; 2000US-0216880.

XX 11-JUL-2000; 2000US-0217487.

XX 11-JUL-2000; 2000US-0217496.

XX 14-JUL-2000; 2000US-0218290.

XX 26-JUL-2000; 2000US-0220963.

XX 26-JUL-2000; 2000US-0220964.

XX 14-AUG-2000; 2000US-0224518.

XX 14-AUG-2000; 2000US-0224519.

XX 14-AUG-2000; 2000US-0225213.

XX 14-AUG-2000; 2000US-0225214.

XX 14-AUG-2000; 2000US-0225266.

XX 14-AUG-2000; 2000US-0225267.

XX 14-AUG-2000; 2000US-0225268.

XX 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.  
 PR 14-AUG-2000; 2000US-0225757.  
 PR 14-AUG-2000; 2000US-0225758.  
 PR 14-AUG-2000; 2000US-0225759.  
 PR 18-AUG-2000; 2000US-0226279.  
 PR 22-AUG-2000; 2000US-0226681.  
 PR 22-AUG-2000; 2000US-0226888.  
 PR 22-AUG-2000; 2000US-0227182.  
 PR 23-AUG-2000; 2000US-0227009.  
 PR 30-AUG-2000; 2000US-0228924.  
 PR 01-SEP-2000; 2000US-0229287.  
 PR 01-SEP-2000; 2000US-0229343.  
 PR 01-SEP-2000; 2000US-0229344.  
 PR 01-SEP-2000; 2000US-0229345.  
 PR 05-SEP-2000; 2000US-0229509.  
 PR 05-SEP-2000; 2000US-0229513.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 06-SEP-2000; 2000US-0230438.  
 PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 08-SEP-2000; 2000US-0232081.  
 PR 12-SEP-2000; 2000US-0231968.  
 PR 14-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 14-SEP-2000; 2000US-0233065.  
 PR 21-SEP-2000; 2000US-0234023.  
 PR 21-SEP-2000; 2000US-0234274.  
 PR 25-SEP-2000; 2000US-0234997.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235835.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 02-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239935.  
 PR 13-OCT-2000; 2000US-0239937.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 20-OCT-2000; 2000US-0242221.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246529.

08-NOV-2000; 2000US-0246532.  
 08-NOV-2000; 2000US-0246609.  
 08-NOV-2000; 2000US-0246610.  
 08-NOV-2000; 2000US-0246611.  
 08-NOV-2000; 2000US-0246613.  
 17-NOV-2000; 2000US-0249207.  
 17-NOV-2000; 2000US-0249208.  
 17-NOV-2000; 2000US-0249209.  
 17-NOV-2000; 2000US-0249210.  
 17-NOV-2000; 2000US-0249211.  
 17-NOV-2000; 2000US-0249212.  
 17-NOV-2000; 2000US-0249213.  
 17-NOV-2000; 2000US-0249214.  
 17-NOV-2000; 2000US-0249215.  
 17-NOV-2000; 2000US-0249216.  
 17-NOV-2000; 2000US-0249217.  
 17-NOV-2000; 2000US-0249218.  
 17-NOV-2000; 2000US-0249244.  
 17-NOV-2000; 2000US-0249245.  
 17-NOV-2000; 2000US-0249264.  
 17-NOV-2000; 2000US-0249265.  
 17-NOV-2000; 2000US-0249297.  
 17-NOV-2000; 2000US-0249299.  
 17-NOV-2000; 2000US-0249300.  
 01-DEC-2000; 2000US-0250391.  
 01-DEC-2000; 2000US-0251160.  
 05-DEC-2000; 2000US-0251030.  
 05-DEC-2000; 2000US-0251988.  
 06-DEC-2000; 2000US-0251719.  
 08-DEC-2000; 2000US-0251479.  
 08-DEC-2000; 2000US-0251856.  
 08-DEC-2000; 2000US-0251868.  
 08-DEC-2000; 2000US-0251869.  
 08-DEC-2000; 2000US-0251989.  
 08-DEC-2000; 2000US-0251990.  
 11-DEC-2000; 2000US-0251990.  
 05-JAN-2001; 2001US-0255678.  
 (HUMA-) HUMAN GENOME SCI INC.  
 Rosen CA, Barash SC, Ruben SM;  
 WPI; 2001-541565/60.  
 N-PSDB; ABA12029.  
 Nucleic acids encoding 3224 human nervous system antigen polypeptides,  
 useful for preventing, diagnosing and/or treating nervous system  
 cancers and metastases -  
 Claim 11; SEQ ID NO 4360; 1701pp + Sequence Listing; English.  
 The invention relates to novel genes (ABA11004-ABA21534) and proteins  
 (AB114678-AB118001) useful for preventing, treating or ameliorating  
 medical conditions e.g. by protein or gene therapy. The genes are  
 isolated from a range of human tissues disclosed in the specification.  
 The nucleic acids, proteins, antibodies and (ant)agonists are useful  
 in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
 and ovarian cancer and other cancers of the adrenal gland, bone, bone  
 marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
 (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
 haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
 disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
 colitis; (c) cardiovascular disorders such as myocardial ischaemia;  
 (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
 epilepsy; and (f) infectious  
 and parasitic infections.  
 Note: The sequence data for this patent did not form part of the  
 printed specification, but was obtained in electronic format directly  
 from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 83 AA;

Query Match 80.5%; Score 33; DB 22; Length 83;

Best Local Similarity 83.3%; Pred. No. 34;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 KHGPRK 7  
 Db 47 RHGPRK 52  
 :|||||  
 RESULT 5  
 AAG27027  
 ID AAG27027 standard; Protein; 150 AA.  
 XX  
 AC AAG27027;  
 XX  
 DT 17-OCT-2000 (first entry)  
 XX  
 DE Zea mays protein fragment SEQ ID NO: 31706.  
 XX  
 KW Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; Genetic mapping; gene expression control; promoter;  
 KW termination sequence; corn.  
 XX  
 OS Zea mays subsp. mays.  
 XX  
 PN EP1033405-A2.  
 XX  
 PD 06-SEP-2000.  
 XX  
 PF 25-FEB-2000; 2000EP-0301439.  
 XX  
 25-FEB-1999; 99US-0121825.  
 PR 05-MAR-1999; 99US-0123180.  
 PR 09-MAR-1999; 99US-0123548.  
 PR 23-MAR-1999; 99US-0125788.  
 PR 25-MAR-1999; 99US-0126264.  
 PR 29-MAR-1999; 99US-0126785.  
 PR 01-APR-1999; 99US-0127462.  
 PR 06-APR-1999; 99US-0128234.  
 PR 08-APR-1999; 99US-0128714.  
 PR 16-APR-1999; 99US-0129845.  
 PR 19-APR-1999; 99US-0130077.  
 PR 21-APR-1999; 99US-0130449.  
 PR 23-APR-1999; 99US-0130510.  
 PR 28-APR-1999; 99US-0130891.  
 PR 30-APR-1999; 99US-0131449.  
 PR 30-APR-1999; 99US-0132048.  
 PR 04-MAY-1999; 99US-0132407.  
 PR 05-MAY-1999; 99US-0132484.  
 PR 05-MAY-1999; 99US-0132485.  
 PR 05-MAY-1999; 99US-0132486.  
 PR 06-MAY-1999; 99US-0132487.  
 PR 07-MAY-1999; 99US-0132863.  
 PR 11-MAY-1999; 99US-0134256.  
 PR 14-MAY-1999; 99US-0134218.  
 PR 14-MAY-1999; 99US-0134219.  
 PR 14-MAY-1999; 99US-0134221.  
 PR 14-MAY-1999; 99US-0134370.  
 PR 18-MAY-1999; 99US-0134768.  
 PR 19-MAY-1999; 99US-0134941.  
 PR 20-MAY-1999; 99US-0135124.  
 PR 21-MAY-1999; 99US-0135353.  
 PR 24-MAY-1999; 99US-0135629.  
 PR 25-MAY-1999; 99US-0136021.  
 PR 27-MAY-1999; 99US-0136392.  
 PR 28-MAY-1999; 99US-0136782.  
 PR 01-JUN-1999; 99US-0137222.  
 PR 03-JUN-1999; 99US-0137528.  
 PR 07-JUN-1999; 99US-0137724.  
 PR 08-JUN-1999; 99US-0138094.  
 PR 10-JUN-1999; 99US-0138540.  
 PR 10-JUN-1999; 99US-0138847.  
 PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139452.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 21-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 23-JUN-1999; 99US-0140695.  
PR 24-JUN-1999; 99US-0140823.  
PR 28-JUN-1999; 99US-0140991.  
PR 29-JUN-1999; 99US-0141287.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 14-JUL-1999; 99US-0144005.  
PR 15-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 16-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145089.  
PR 23-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.

PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 17-AUG-1999; 99US-0149426.  
PR 18-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149721.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0158293.  
PR 13-OCT-1999; 99US-0158294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 80.5%; Score 33; DB 21; Length 150;  
Best Local Similarity 83.3%; Pred. No. 61;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TRKGPR 6  
Db 69 TRKGPR 74